Phase 2.0

Quick Start Guide

Schrödinger Press

Revision A, April 2006

Contents

Document Conventions

In addition to the use of italics for names of documents, the font conventions that are used in this document are summarized in the table below.

Table 3.1.

Font	Example	Use
Sans serif	Project Table	Names of GUI features, such as panels, menus, menu items, buttons, and labels
Monospace	`$SCHRODINGER/maestro`	File names, directory names, commands, environment variables, and screen output
Italic	*filename*	Text that the user must replace with a value
Sans serif uppercase	CTRL+H	Keyboard keys

In descriptions of command syntax, the following UNIX conventions are used: braces `{ }` enclose a choice of required items, square brackets `[ ]` enclose optional items, and the bar symbol `|` separates items in a list from which one item must be chosen. Lines of command syntax that wrap should be interpreted as a single command.

In this document, to *type* text means to type the required text in the specified location, and to *enter* text means to type the required text, then press the ENTER key.

References to literature sources are given in square brackets, like this: [10].

Overview

1.1 About Phase

Phase is a versatile product for pharmacophore perception, structure alignment, activity prediction, and 3D database searching. Given a set of molecules with high affinity for a particular protein target, Phase utilizes fine-grained conformational sampling and a range of scoring techniques to identify common pharmacophore hypotheses, which convey characteristics of 3D chemical structures that are purported to be critical for binding. Each hypothesis is accompanied by a set of aligned conformations that suggest the relative manner in which the molecules are likely to bind.

A given hypothesis may be combined with known activity data to create 3D QSAR models that identify overall aspects of molecular structure that govern activity. These models may be used in conjunction with the hypothesis to mine a 3D database for molecules that are most likely to exhibit strong activity toward the target.

Phase provides support for lead discovery, SAR development, lead optimization and lead expansion. Phase may also be used as a source of molecular alignments for third-party 3D QSAR programs.

Phase is integrated into Maestro, the graphical user interface (GUI) for all Schrödinger products. An introduction to the general capabilities of Maestro is given in Chapter 2. For more detailed information on Maestro, see the Maestro online help, the *Maestro User Manual*, or the *Maestro Tutorial*. For detailed information on Phase, see the *Phase User Manual*.

1.2 About this Document

This document provides tutorial instruction in the three main Phase workflows.

- Chapter 3 contains exercises on developing a pharmacophore model and building QSAR models.
- Chapter 4 contains exercises on preparing a database for searching. The database will contain both single structures, without conformers and sites, and structures for which the conformers and sites are generated and stored in the database.

- Chapter 5 contains exercises on searching the database for matches to a hypothesis, using the database prepared in the exercises in Chapter 4, and a hypothesis supplied with the distribution. You do not need to complete the exercises in Chapter 3 to work through this chapter.

1.3 Command-Line Tutorials

In addition to the exercises for running Phase programs from Maestro, a set of tutorials for running Phase programs from the command line is provided with the software distribution. Each tutorial consists of a PDF file containing the instructions and a gzipped archive (`.tar`) file containing the necessary files. The available tutorials are listed in Table 1.1.

Table 1.1. List of command-line tutorials

Tutorial	Documentation	Archive file
Pharmacophore Model Development	`pharm_tutorial.pdf`	`pharm_tutorial.tar.gz`
Database Management and Searching	`db_tutorial.pdf`	`db_tutorial.tar.gz`
Searching for Matches on a Grid	`gridSearch_tutorial.pdf`	`gridSearch_tutorial.tar.gz`

Introduction to Maestro

Maestro is the graphical user interface for all of Schrödinger's products: CombiGlide™, Epik™, Glide™, Impact™, Jaguar™, Liaison™, LigPrep™, MacroModel®, Phase™, Prime™, QikProp™, QSite™, SiteMap™, and Strike™. It contains tools for building, displaying, and manipulating chemical structures; for organizing, loading, and storing these structures and associated data; and for setting up, monitoring, and visualizing the results of calculations on these structures. This chapter provides a brief introduction to Maestro and some of its capabilities. For more information on any of the topics in this chapter, see the *Maestro User Manual*.

2.1 General Interface Behavior

Most Maestro panels are amodal: more than one panel can be open at a time, and a panel need not be closed for an action to be carried out. Each Maestro panel has a Close button so you can hide the panel from view.

Maestro supports the mouse functions common to many graphical user interfaces. The left button is used for choosing menu items, clicking buttons, and selecting objects by clicking or dragging. This button is also used for resizing and moving panels. The right button displays a shortcut menu. Other common mouse functions are supported, such as using the mouse in combination with the SHIFT or CTRL keys to select a range of items and select or deselect a single item without affecting other items.

In addition, the mouse buttons are used for special functions described later in this chapter. These functions assume that you have a three-button mouse. If you have a two-button mouse, ensure that it is configured for three-button mouse simulation (the middle mouse button is simulated by pressing or holding down both buttons simultaneously).

2.2 Starting Maestro

Before starting Maestro, you must first set the `SCHRODINGER` environment variable to point to the installation directory. To set this variable, enter the following command at a shell prompt:

csh/tcsh: `setenv SCHRODINGER` *installation-directory*

bash/ksh: `export SCHRODINGER=`*installation-directory*

You might also need to set the `DISPLAY` environment variable, if it is not set automatically when you log in. To determine if you need to set this variable, enter the command:

```
echo $DISPLAY
```

If the response is a blank line, set the variable by entering the following command:

csh/tcsh: `setenv DISPLAY` *display-machine-name*`:0.0`

bash/ksh: `export DISPLAY=`*display-machine-name*`:0.0`

After you set the `SCHRODINGER` and `DISPLAY` environment variables, you can start Maestro using the command:

`$SCHRODINGER/maestro` *options*

If you add the `$SCHRODINGER` directory to your path, you only need to enter the command `maestro`. Options for this command are given in Section 2.1 of the *Maestro User Manual*.

The directory from which you started Maestro is Maestro's current working directory, and all data files are written to and read from this directory unless otherwise specified (see Section 2.8 on page 25). You can change directories by entering the following command in the command input area (see page 6) of the main window:

`cd` *directory-name*

where *directory-name* is either a full path or a relative path.

2.3 The Maestro Main Window

The Maestro main window is shown in Figure 2.1 on page 5. The main window components are listed below.

The following components are always visible:

- **Title bar**—displays the Maestro version, the project name (if there is one) and the current working directory.
- **Auto-Help**—automatically displays context-sensitive help.
- **Menu bar**—provides access to panels.
- **Workspace**—displays molecular structures and other 3D graphical objects.

The following components can be displayed or hidden by choosing the component from the Display menu. Your choice of which main window components are displayed is persistent between Maestro sessions.

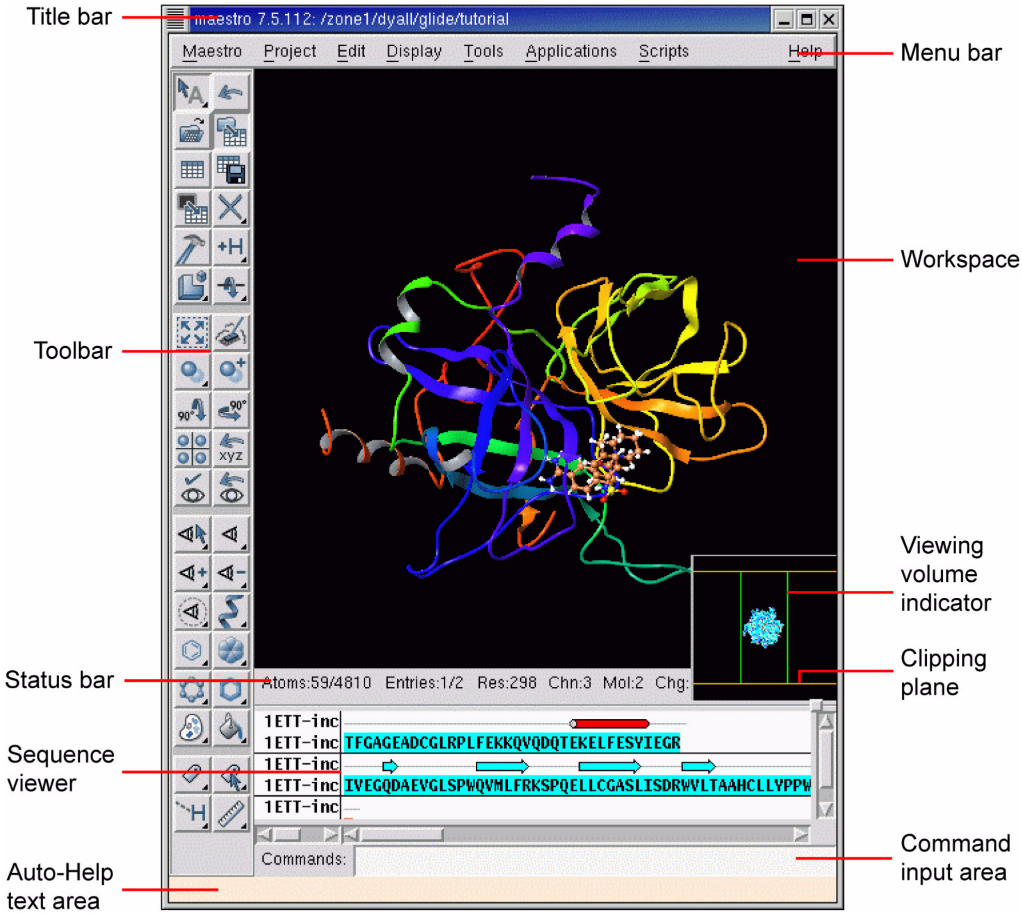


Figure 2.1. The Maestro main window.

- **Toolbar**—contains buttons for many common tasks and provides tools for displaying and manipulating structures, as well as organizing the Workspace.
- **Status bar**—displays information about a particular atom, or about structures in the Workspace, depending on where the pointer pauses (see Section 2.5 of the *Maestro User Manual* for details):
  - **Atom**—displays the chain, residue number, element, PDB atom name, formal charge, and title or entry name (this last field is set by choosing Preferences from the Maestro menu and selecting the Feedback folder).
  - **Workspace**—displays the number of atoms, entries, residues, chains, and molecules in the Workspace.

- **Clipping planes window**—displays a small, top view of the Workspace and shows the clipping planes and viewing volume indicators.
- **Sequence viewer**—shows the sequences for proteins displayed in the Workspace. See Section 2.6 of the *Maestro User Manual* for details.
- **Command input area**—provides a place to enter Maestro commands.

When a distinction between components in the main window and those in other panels is needed, the term *main* is applied to the main window components (e.g., main toolbar).

You can expand the Workspace to occupy the full screen, by pressing CTRL+=. All other components and panels are hidden. To return to the previous display, press CTRL+= again.

2.3.1 The Menu Bar

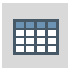
The menus on the main menu bar provide access to panels, allow you to execute commands, and control the appearance of the Workspace. The main menus are as follows:

- Maestro—save or print images in the Workspace, execute system commands, save or load a panel layout, set preferences, set up Maestro command aliases, and quit Maestro.
- Project—open and close projects, import and export structures, make a snapshot, and annotate a project. These actions can also be performed from the Project Table panel. For more information, see Section 2.4 on page 11.
- Edit—undo actions, build and modify structures, define command scripts and macros, and find atoms in the Workspace.
- Display—control the display of the contents of the Workspace, arrange panels, and display or hide main window components.
- Tools—group atoms; measure, align, and superimpose structures; and view and visualize data.
- Applications—set up, submit, and monitor jobs for Schrödinger's computational programs. Some products have a submenu from which you can choose the task to be performed.
- Scripts—manage and install Python scripts that come with the distribution and scripts that you create yourself. (See Chapter 13 of the *Maestro User Manual* for details.)
- Help—open the Help panel, the PDF documentation index, or information panels; run a demonstration; and display or hide Balloon Help (tooltips).

2.3.2 The Toolbar

The main toolbar contains three kinds of buttons for performing common tasks:

Action—Perform a simple task, like clearing the Workspace.



Display—Open or close a panel or open a dialog box, such as the Project Table panel.

Menu—Display a *button menu*. These buttons have a triangle in the lower right corner.

There are four types of items on button menus, and all four types can be on the same menu (see Figure 2.2):

- **Action**—Perform an action immediately.
- **Display**—Open a panel or dialog box.
- **Object types for selection**—Choose Atoms, Bonds, Residues, Chains, Molecules, or Entries, then click on an atom in the Workspace to perform the action on all the atoms in that structural unit.

  The object type is marked on the menu with a red diamond and the button is indented to indicate the action to be performed.
- **Other setting**—Set a state, choose an attribute, or choose a parameter and click on atoms in the Workspace to display or change that parameter.

The toolbar buttons are described below. Some descriptions refer to features not described in this chapter. See the *Maestro User Manual* for a fuller description of these features.

Figure 2.2. The Workspace selection ***button menu and the*** Adjust distances, angles or dihedrals ***button menu.***

Workspace selection
– Choose an object type for selecting
– Open the Atom Selection dialog box

Undo/Redo
Undo or redo the last action. Performs the same function as the Undo item on the Edit menu, and changes to an arrow pointing in the opposite direction when an Undo has been performed, indicating that its next action is Redo.

Open a project
Open the Open Project dialog box.

Import structures
Open the Import panel.

Open/Close Project Table
Open the Project Table panel or close it if it is open.

Save as
Open the Save Project As dialog box, to save the project with a new name.

Create entry from Workspace
Open a dialog box in which you can create an entry in the current project using the contents of the Workspace.

Delete
– Choose an object type for deletion
– Delete hydrogens and waters
– Open the Atom Selection dialog box
– Delete other items associated with the structures in the Workspace
– Click to select atoms to delete
– Double-click to delete all atoms

Open/Close Build panel
Open the Build panel or close it if it is open.

Add hydrogens
– Choose an object type for applying a hydrogen treatment
– Open the Atom Selection dialog box
– Click to select atoms to treat
– Double-click to apply to all atoms

Local transformation
– Choose an object type for transforming
– Click to select atoms to transform
– Open the Advanced Transformations panel

Adjust distances, angles or dihedrals
– Choose a parameter for adjusting
– Delete adjustments

Fit to screen
Scale the displayed structure to fit into the Workspace and reset the center of rotation.

Clear Workspace
Clear all atoms from the Workspace.

Set fog display state
Choose a fog state. Automatic means fog is on when there are more than 40 atoms in the Workspace, otherwise it is off.

Enhance depth cues
Optimize fogging and other depth cues based on what is in the Workspace.

Rotate around X axis by 90 degrees
Rotate the Workspace contents around the X axis by 90 degrees.

Rotate around Y axis by 90 degrees
Rotate the Workspace contents around the Y axis by 90 degrees.

Tile entries
Arrange entries in a rectangular grid in the Workspace.

Reset Workspace
Reset the rotation, translation, and zoom of the Workspace to the default state.

Save view
Save the current view of the Workspace: orientation, location, and zoom.

Restore view
Restore the last saved view of the Workspace: orientation, location, and zoom.

Display only selected atoms
– Choose an object type for displaying
– Click to select atoms to display
– Double-click to display all atoms

Display only
– Choose a predefined atom category
– Open the Atom Selection dialog box

Also display
– Choose a predefined atom category
– Open the Atom Selection dialog box

Undisplay
– Choose a predefined atom category
– Open the Atom Selection dialog box

Display residues within N angstroms of currently displayed atoms
– Choose a radius
– Open a dialog box to set a value

Show, hide, or color ribbons
– Choose to show or hide ribbons
– Choose a color scheme for coloring ribbons

Draw bonds in wire
– Choose an object type for drawing bonds in wire representation
– Open the Atom Selection dialog box
– Click to select atoms for representation
– Double-click to apply to all atoms

Draw atoms in CPK
– Choose an object type for drawing bonds in CPK representation
– Open the Atom Selection dialog box
– Click to select atoms for representation
– Double-click to apply to all atoms

Draw atoms in Ball & Stick
– Choose an object type for drawing bonds in Ball & Stick representation
– Open the Atom Selection dialog box
– Click to select atoms for representation
– Double-click to apply to all atoms

Draw bonds in tube
– Choose an object type for drawing bonds in tube representation
– Open the Atom Selection dialog box
– Click to select atoms for representation
– Double-click to apply to all atoms

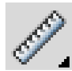
Color all atoms by scheme
Choose a predefined color scheme.

Color residue by constant color
– Choose a color for applying to residues
– Click to select residues to color
– Double-click to color all atoms

Label atoms
– Choose a predefined label type
– Delete labels

Label picked atoms
– Choose an object type for labeling atoms
– Open the Atom Selection dialog box
– Open the Atom Labels panel at the Composition folder
– Delete labels
– Click to select atoms to label
– Double-click to label all atoms

Display H-bonds
– Choose bond type:
intra—displays H-bonds within the selected molecule
inter—displays H-bonds between the selected molecule and all other atoms.
– Delete H-bonds
– Click to select molecule



Measure distances, angles or dihedrals
– Choose a parameter for displaying measurements
– Delete measurements
– Click to select atoms for measurement

2.3.3 Mouse Functions in the Workspace

The left mouse button is used for selecting objects. You can either click on a single atom or bond, or you can drag to select multiple objects. The right mouse button opens shortcut menus, which are described in Section 2.7 of the *Maestro User Manual*.

The middle and right mouse buttons can be used on their own and in combination with the SHIFT and CTRL keys to perform common operations, such as rotating, translating, centering, adjusting, and zooming.

Table 2.1. Mapping of Workspace operations to mouse actions.

Mouse Button	Keyboard	Motion	Action
Left		click, drag	Select
Left	SHIFT	click, drag	Toggle the selection
Middle		drag	Rotate about X and Y axes Adjust bond, angle, or dihedral
Middle	SHIFT	drag vertically	Rotate about X axis
Middle	SHIFT	drag horizontally	Rotate about Y axis
Middle	CTRL	drag horizontally	Rotate about Z axis
Middle	SHIFT + CTRL	drag horizontally	Zoom
Right		click	Spot-center on selection
Right		click and hold	Display shortcut menu
Right		drag	Translate in the X-Y plane
Right	SHIFT	drag vertically	Translate along the X axis
Right	SHIFT	drag horizontally	Translate along the Y axis
Right	CTRL	drag horizontally	Translate along the Z axis
Middle & Right		drag horizontally	Zoom

2.3.4 Shortcut Key Combinations

Some frequently used operations have been assigned shortcut key combinations. The shortcuts available in the main window are described in Table 2.2.

Table 2.2. Shortcut keys in the Maestro main window.

Keys	Action	Equivalent Menu Choices
CTRL+B	Open Build panel	Edit > Build
CTRL+C	Create entry	Project > Create Entry From Workspace
CTRL+E	Open Command Script Editor panel	Edit > Command Script Editor
CTRL+F	Open Find Atoms panel	Edit > Find
CTRL+H	Open Help panel	Help > Help
CTRL+I	Open Import panel	Project > Import Structures
CTRL+M	Open Measurements panel	Tools > Measurements
CTRL+N	Create new project	Project > New
CTRL+O	Open project	Project > Open
CTRL+P	Print	Maestro > Print
CTRL+Q	Quit	Maestro > Quit
CTRL+S	Open Sets panel	Tools > Sets
CTRL+T	Open Project Table panel	Project > Show Table
CTRL+W	Close project	Project > Close
CTRL+Z	Undo/Redo last command	Edit > Undo/Redo
CTRL+=	Enter and exit full screen mode (Workspace occupies full screen)	None

2.4 Maestro Projects

All the work you do in Maestro is done within a *project*. A project consists of a set of *entries*, each of which contains one or more chemical structures and their associated data. In any Maestro session, there can be only one Maestro project open. If you do not specify a project when you start Maestro, a *scratch* project is created. You can work in a scratch project without saving it, but you must save it in order to use it in future sessions. When you save or close a project, all the view transformations (rotation, translation, and zoom) are saved with it. When you close a project, a new scratch project is automatically created.

Likewise, if there is no entry displayed in the Workspace, Maestro creates a *scratch* entry. Structures that you build in the Workspace constitute a scratch entry until you save the structures as project entries. The scratch entry is not saved with the project unless you explicitly add it to the project. However, you can use a scratch entry as input for some calculations.

To add a scratch entry to a project, do one of the following:

- Click the Create entry from Workspace button:

- Choose Create Entry from Workspace from the Project menu.
- Press CTRL+C.

In the dialog box, enter a name and a title for the entry. The entry name is used internally to identify the entry and can be modified by Maestro. The title can be set or changed by the user, but is not otherwise modified by Maestro.

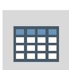
Once an entry has been incorporated into the project, its structures and their data are represented by a row in the Project Table. Each row contains the row number, an icon indicating whether the entry is displayed in the Workspace (the In column), the entry title, a button to open the Surfaces panel if the entry has surfaces, the entry name, and any entry properties. The row number is not a property of the entry.

Entries can be collected into groups, and the members of the group can be displayed or hidden. Most additions of multiple entries to the Project Table are done as entry groups.

You can use entries as input for all of the computational programs—Glide, Impact, Jaguar, Liaison, LigPrep, MacroModel, Phase, Prime, QikProp, QSite, and Strike. You can select entries as input for the ePlayer, which displays the selected structures in sequence. You can also duplicate, combine, rename, and sort entries; create properties; import structures as entries; and export structures and properties from entries in various formats.

To open the Project Table panel, do one of the following:

- Click the Open/Close Project Table button on the toolbar



- Choose Show Table from the Project menu
- Press CTRL+T.

The Project Table panel contains a menu bar, a toolbar, and the table itself.

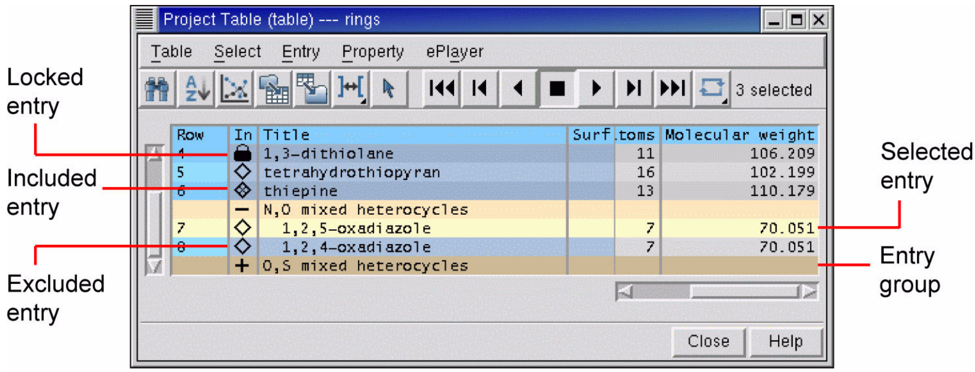
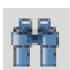


Figure 2.3. The* Project Table *panel.

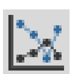
2.4.1 The Project Table Toolbar

The Project Table toolbar contains two groups of buttons and a status display. The first set of buttons opens various panels that allow you to perform functions on the entries in the Project Table. The second set of buttons controls the ePlayer, which "plays through" the selected structures: each structure is displayed in the Workspace in sequence, at a given time interval. See Section 2.3.2 on page 7 for a description of the types of toolbar buttons. The buttons are described below.



Find
Open the Find panel for locating alphanumeric text in any column of the Project Table, except for the row number.

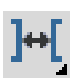
Sort
Open the Sort panel for sorting entries by up to three properties.



Plot
Open the Plot panel for plotting entry properties.

Import Structure
Open the Import panel for importing structures into the project.

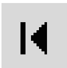
Export Structure
Open the Export panel for exporting structures to a file.



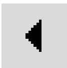
Columns
Choose an option for adjusting the column widths.

Select only
Open the Entry Selection dialog box for selecting entries based on criteria for entry properties.

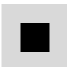
Go to start
Display the first selected structure.



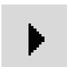
Previous
Display the previous structure in the list of selected structures.



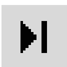
Play backward
Display the selected structures in sequence, moving toward the first.



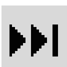
Stop
Stop the ePlayer.



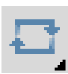
Play forward
Display the selected structures in sequence, moving toward the last.



Next
Display the next structure in the list of selected structures.



Go to end
Display the last selected structure.



Loop
Choose an option for repeating the display of the structures. Single Direction displays structures in a single direction, then repeats. Oscillate reverses direction each time the beginning or end of the list is reached.

The status display, to the right of the toolbar buttons, shows the number of selected entries. When you pause the cursor over the status display, the Balloon Help shows the total number of entries, the number shown in the table, the number selected, and the number included in the Workspace.

2.4.2 The Project Table Menus

- Table—find text, sort entries, plot properties, import and export structures, and configure the Project Table.
- Select—select all entries, none, invert your selection, or select classes of entries using the Entry Selection dialog box and the Filter panel.
- Entry—include or exclude entries from the Workspace, display or hide entries in the Project Table, and perform various operations on the selected entries.
- Property—display and manipulate entry properties in the Project Table.
- ePlayer—view entries in succession, stop, reverse, and set the ePlayer options.

2.4.3 Selecting Entries

Many operations in Maestro are performed on the entries selected in the Project Table. The Project Table functions much like any other table: select rows by clicking, shift-clicking, and control-clicking. However, because clicking in an editable cell of a selected row enters edit mode, you should click in the Row column to select entries. See Section 2.4.5 on page 16 for more information on mouse actions in the Project Table. There are shortcuts for selecting classes of entries on the Select menu.

In addition to selecting entries manually, you can select entries that meet a combination of conditions on their properties. Such combinations of conditions are called *filters*. Filters are Entry Selection Language (ESL) expressions and are evaluated at the time they are applied. For example, if you want to set up a Glide job that uses ligands with a low molecular weight (say, less than 300) and that has certain QikProp properties, you can set up a filter and use it to select entries for the job. If you save the filter, you can use it again on a different set of ligands that meet the same selection criteria.

To create a filter:

1. Do one of the following:
   - Choose Only, Add, or Deselect from the Select menu.
   - Click the Entry selection button on the toolbar.

2. In the Properties folder, select a property from the property list, then select a condition.
3. Combine this selection with the current filter by clicking Add, Subtract, or Intersect. These buttons perform the Boolean operations OR, AND NOT, and AND on the corresponding ESL expressions.
4. To save the filter for future use click Create Filter, enter a name, and click OK.
5. Click OK to apply the filter immediately.

2.4.4 Including Entries in the Workspace

In addition to selecting entries, you can also use the Project Table to control which entries are displayed in the Workspace. An entry that is displayed in the Workspace is *included* in the Workspace; likewise, an entry that is not displayed is *excluded*. Included entries are marked by an X in the diamond in the In column; excluded entries are marked by an empty diamond. Entry inclusion is completely independent of entry selection.

To include or exclude entries, click, shift-click, or control-click in the In column of the entries, or select entries and choose Include or Exclude from the Entry menu. Inclusion with the mouse works just like selection: when you include an entry by clicking, all other entries are excluded.

It is sometimes useful to keep one entry in the Workspace and include others one by one: for example, a receptor and a set of ligands. You can fix the receptor in the Workspace by selecting it in the Project Table and choosing Fix from the Entry menu or by pressing CTRL+F. A padlock icon replaces the diamond in the In column to denote a *fixed* entry. To remove a fixed entry from the Workspace, you must exclude it explicitly (CTRL+X). It is not affected by the inclusion or exclusion of other entries. Fixing an entry affects only its inclusion; you can still rotate, translate, or modify the structure.

2.4.5 Mouse Functions in the Project Table

The Project Table supports the standard use of shift-click and control-click to select objects. This behavior applies to the selection of entries and the inclusion of entries in the Workspace. You can also drag to resize rows and columns and to move rows.

You can drag a set of non-contiguous entries to reposition them in the Project Table. When you release the mouse button, the entries are placed after the first unselected entry that precedes the entry on which the cursor is resting. For example, if you select entries 2, 4, and 6, and release the mouse button on entry 3, these three entries are placed after entry 1, because entry 1 is the first unselected entry that precedes entry 3. To move entries to the top of the table, drag them above the top of the table; to move entries to the end of the table, drag them below the end of the table.

A summary of mouse functions in the Project Table is provided in Table 2.3.

Table 2.3. Mouse operations in the Project Table.

Task	Mouse Operation
Change a Boolean property value	Click repeatedly in a cell to cycle through the possible values (On, Off, Clear)
Display the Entry menu for an entry	Right-click anywhere in the entry. If the entry is not selected, it becomes the selected entry. If the entry is selected, the action is applied to all selected entries.
Display a version of the Property menu for a property	Right-click in the column header
Edit the text or the value in a table cell	Click in the cell and edit the text or value
Include an entry in the Workspace, exclude all others	Click the In column of the entry

Table 2.3. Mouse operations in the Project Table. (Continued)

Task	Mouse Operation
Move selected entries	Drag the entries
Paste text into a table cell	Middle-click
Resize rows or columns	Drag the boundary with the middle mouse button
Select an entry, deselect all others	For an unselected entry, click anywhere in the row except the In column; for a selected entry, click the row number.
Select or include multiple entries	Click the first entry then shift-click the last entry
Toggle the selection or inclusion state	Control-click the entry or the In column

2.4.6 Project Table Shortcut Keys

Some frequently used project operations have been assigned shortcut key combinations. The shortcuts, their functions, and their menu equivalents are listed in Table 2.4.

Table 2.4. Shortcut keys in the Project Table.

Keys	Action	Equivalent Menu Choices
CTRL+A	Select all entries	Select > All
CTRL+F	Fix entry in Workspace	Entry > Fix
CTRL+I	Open Import panel	Table > Import Structures
CTRL+N	Include only selected entries	Entry > Include Only
CTRL+U	Deselect all entries	Select > None
CTRL+X	Exclude selected entries	Entry > Exclude
CTRL+Z	Undo/Redo last command	Edit > Undo/Redo in main window

2.5 Building a Structure

After you start Maestro, the first task is usually to create or import a structure. You can open existing Maestro projects or import structures from other sources to obtain a structure, or you can build your own. To open the Build panel, do one of the following:

- Click the Open/Close Build panel button in the toolbar:

- Choose Build from the Edit menu.
- Press CTRL+B.

The Build panel allows you to create structures by drawing or placing atoms or fragments in the Workspace and connecting them into a larger structure, to adjust atom positions and bond orders, and to change atom properties. This panel contains a toolbar and three folders.

2.5.1 Placing and Connecting Fragments

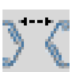
The Build panel provides several tools for creating structures in the Workspace. You can place and connect fragments, or you can draw a structure freehand.

To place a fragment in the Workspace:

1. Select Place.
2. Choose a fragment library from the Fragments menu.
3. Click a fragment.
4. Click in the Workspace where you want the fragment to be placed.

To connect fragments in the Workspace, do one of the following:

- Place another fragment and connect them using the Connect & Fuse panel, which you open from the Edit menu on the main menu bar or with the Display Connect & Fuse panel on the Build toolbar.

- Replace one or more atoms in the existing fragment with another fragment by selecting a fragment and clicking in the Workspace on the main atom to be replaced.
- Grow another fragment by selecting Grow in the Build panel and clicking the fragment you want to add in the Fragments folder.

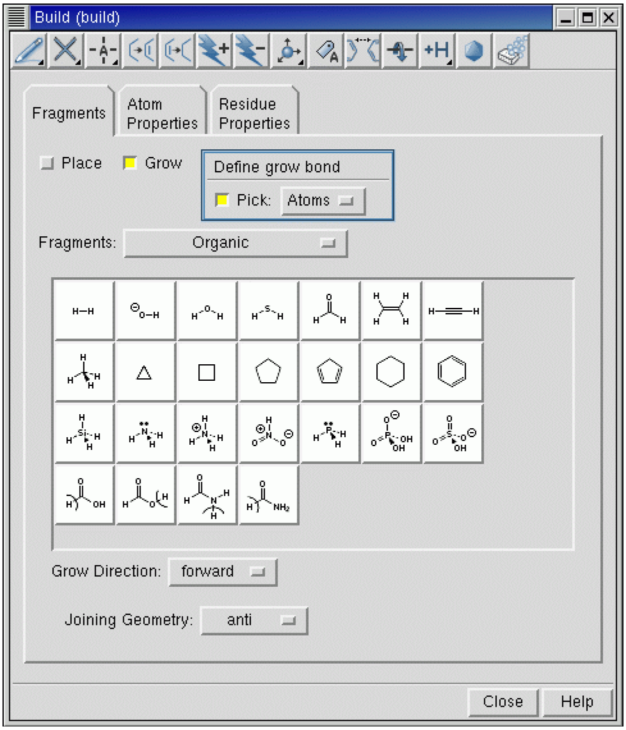
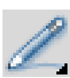


Figure 2.4. The* Build *panel.

Grow mode uses predefined rules to connect a fragment to the *grow bond*. The grow bond is marked by a green arrow. The new fragment replaces the atom at the head of the arrow on the grow bond and all atoms attached to it. To change the grow bond, choose Bonds from the Pick option menu in the Build panel and click on the desired grow bond in the Workspace. The arrow points to the atom nearest to where you clicked.

To draw a structure freehand:

1. Choose an element from the Draw button menu on the Build panel toolbar:



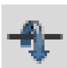
2. Click in the Workspace to place an atom of that element.
3. Click again to place another atom and connect it to the previous atom.
4. Continue this process until you have drawn the structure.
5. Click the active atom again to finish drawing.

2.5.2 Adjusting Properties

In the Atom Properties folder, you can change the properties of the atoms in the Workspace. For each item on the Property option menu—Element, Atom Type (MacroModel), Partial Charge, PDB Atom Name, Grow Name, and Atom Name—there is a set of tools you can use to change the atom properties. For example, the Element tools consist of a periodic table from which you can choose an element and select an atom to change it to an atom of the selected element.

Similarly, the Residue Properties folder provides tools for changing the properties of residues: the Residue Number, the Residue Name, and the Chain Name.

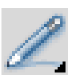
To adjust bond lengths, bond angles, dihedral angles, and chiralities during or after building a structure, use the Adjust distances, angles or dihedrals button on the main toolbar:



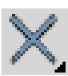
You can also open the Adjust panel from this button menu, from the Display Adjust panel button on the Build panel toolbar (which has the same appearance as the above button) or from the Edit menu in the main window.

2.5.3 The Build Panel Toolbar

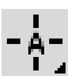
The toolbar of the Build panel provides quick access to tools for drawing and modifying structures and labeling atoms. See Section 2.3.2 on page 7 for a description of the types of toolbar buttons. The toolbar buttons and their use are described below.



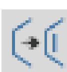
Free-hand drawing
Choose an element for drawing structures freehand in the Workspace (default C). Each click in the Workspace places an atom and connects it to the previous atom.



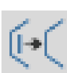
Delete
Choose an object for deleting. Same as the Delete button on the main toolbar, see page 8.



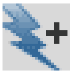
Set element
Choose an element for changing atoms in the Workspace (default C). Click an atom to change it to the selected element.



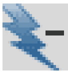
Increment bond order
Select a bond to increase its bond order by one, to a maximum of 3.



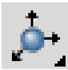
Decrement bond order
Select a bond to decrease its bond order by one, to a minimum of 0.



Increment formal charge
Select an atom to increase its formal charge by one.

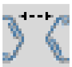


Decrement formal charge
Select an atom to decrease its formal charge by one.

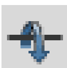


Move
Choose a direction for moving atoms, then click the atom to be moved. Moves in the XY plane are made by clicking the new location. Moves in the Z direction are made in 0.5 Å increments.

Label
Apply heteroatom labels as you build a structure. The label consists of the element name and formal charge, and is applied to atoms other than C and H.



Display Connect & Fuse panel
Open the Connect & Fuse panel so you can connect structures (create bonds between structures) or fuse structures (replace atoms of one structure with those of another).



Display Adjust panel
Open the Adjust panel so you can change bond lengths, bond angles, dihedral angles, or atom chiralities.

Add hydrogens
Choose an atom type for applying the current hydrogen treatment. Same as the Add hydrogens button on the main toolbar, see page 8.

Geometry Symmetrizer
Open the Geometry Symmetrizer panel for symmetrizing the geometry of the structure in the Workspace.

Geometry Cleanup
Clean up the geometry of the structure in the Workspace.

2.6 Selecting Atoms

Maestro has a powerful set of tools for selecting atoms in a structure: toolbar buttons, picking tools in panels, and the Atom Selection dialog box. These tools allow you to select atoms in two ways:

- Select atoms first and apply an action to them
- Choose an action first and then select atoms for that action

2.6.1 Toolbar Buttons

The small triangle in the lower right corner of a toolbar button indicates that the button contains a menu. Many of these buttons allow you to choose an object type for selecting: choose Atoms, Bonds, Residues, Chains, Molecules, or Entries, then click on an atom in the Workspace to perform the action on all the atoms in that structural unit.

For example, to select atoms with the Workspace selection toolbar button:

1. Choose Residues from the Workspace selection button menu:

   The button changes to:

2. Click on an atom in a residue in the Workspace to select all the atoms in that residue.

2.6.2 Picking Tools

The picking tools are embedded in each panel in which you need to select atoms to apply an operation. The picking tools in a panel can include one or more of the following:

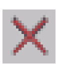
- Pick option menu—Allows you to choose an object type. Depending on the operation to be performed, you can choose Atoms, Bonds, Residues, Chains, Molecules, or Entries, then click on an atom in the Workspace to perform the action on all the atoms in that structural unit.

  The Pick option menu varies from panel to panel, because not all object types are appropriate for a given operation. For example, some panels have only Atoms and Bonds in the Pick option menu.

- All button—Performs the action on all atoms in the Workspace.
- Selection button—Performs the action on any atoms already selected in the Workspace.
- Previous button—Performs the action on the most recent atom selection defined in the Atom Selection dialog box.
- Select button—Opens the Atom Selection dialog box.
- ASL text box—Allows you to type in an ASL expression for selecting atoms.

  ASL stands for Atom Specification Language, and is described in detail in the *Maestro Command Reference Manual*.

- Clear button—Clears the current selection

- Show markers option—Marks the selected atoms in the Workspace.

For example, to label atoms with the Label Atoms panel:

1. Choose Atom Labels from the Display menu.
2. In the Composition folder, select Element and Atom Number.
3. In the picking tools section at the top of the panel, you could do one of the following:
    - Click Selection to apply labels to the atoms already selected in the Workspace (from the previous example).
    - Choose Residues from the Pick option menu and click on an atom in a different residue to label all the atoms in that residue.

2.6.3 The Atom Selection Dialog Box

If you wish to select atoms based on more complex criteria, you can use the Atom Selection dialog box. To open this dialog box, choose Select from a button menu or click the Select button in a panel. See Section 5.3 of the *Maestro User Manual* for detailed instructions on how to use the Atom Selection dialog box.

2.7 Scripting in Maestro

Although you can perform nearly all Maestro-supported operations through menus and panels, you can also perform operations using Maestro commands, or compilations of these commands, called *scripts*. Scripts can be used to automate lengthy procedures or repetitive tasks and can be created in several ways. These are summarized below.

2.7.1 Python Scripts

Python is a full-featured scripting language that has been embedded in Maestro to extend its scripting facilities. The Python capabilities within Maestro include access to Maestro functionality for dealing with chemical structures, projects, and Maestro files.

The two main Python commands used in Maestro are:

- `pythonrun`—executes a Python module. (You can also use the alias `pyrun`.) The syntax is:

    `pythonrun` *module*.*function*

- `pythonimport`—rereads a Python file so that the next time you use the `pythonrun` command, it uses the updated version of the module. (You can also use the alias `pyimp`.)

From the Maestro Scripts menu you can install, manage, and run Python scripts. For more information on the Scripts menu, see Section 13.1 of the *Maestro User Manual*.

For more information on using Python with Maestro, see *Scripting with Python*.

2.7.2 Command Scripts

All Maestro commands are logged and displayed in the Command Script Editor panel. This means you can create a command script by performing the operations with the GUI controls, copying the logged commands from the Command History list into the Script text area of the panel, then saving the list of copied commands as a script.

To run an existing command script:

1. Open the Command Script Editor panel from the Edit menu in the main window.
2. Click Open Local and navigate to the directory containing the desired script.
3. Select a script in the Files list and click Open.

   The script is loaded into the Script window of the Command Script Editor panel.
4. Click Run Script.

Command scripts cannot be used for Prime operations.

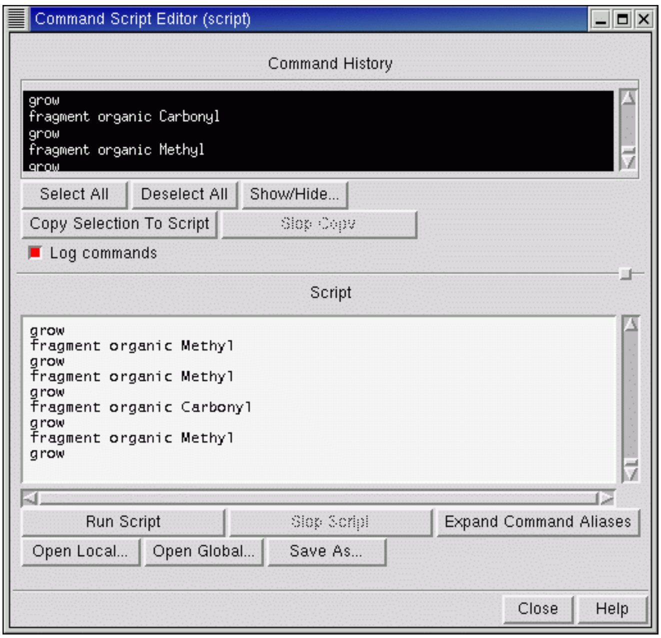


Figure 2.5. The Command Script Editor ***panel.***

2.7.3 Macros

There are two kinds of macros you can create: named macros and macros assigned to function keys F1 through F12.

To create and run a named macro:

1. Open the Macros panel from the Edit menu in the main window.
2. Click New, enter a name for the macro, and click OK.
3. In the Definition text box, type the commands for the macro.
4. Click Update to update the macro definition.
5. To run the macro, enter the following in the command input area in the main window:

   `macrorun` *macro-name*

   If the command input area is not visible, choose Command Input Area from the Display menu.

To create and run a function key macro:

1. Open the Function Key Macros panel from the Edit menu in the main window.
2. From the Macro Key option, select a function key (F1 through F12) to which to assign the macro.
3. In the text box, type the commands for the macro.
4. Click Run to test the macro or click Save to save it.
5. To run the macro from the main window, press the assigned function key.

For more information on macros, see Section 13.5 of the *Maestro User Manual*.

2.8 Specifying a Maestro Working Directory

When you use Maestro to launch Phase jobs, Maestro writes job output to the directory specified in the Directory folder of the Preferences panel. By default, this directory (the file I/O directory) is the directory from which you started Maestro.

To change the Maestro working directory:

1. Open the Preferences panel from the Maestro menu.
2. Click the Directory tab.
3. Select the directory you want to use for reading and writing files.

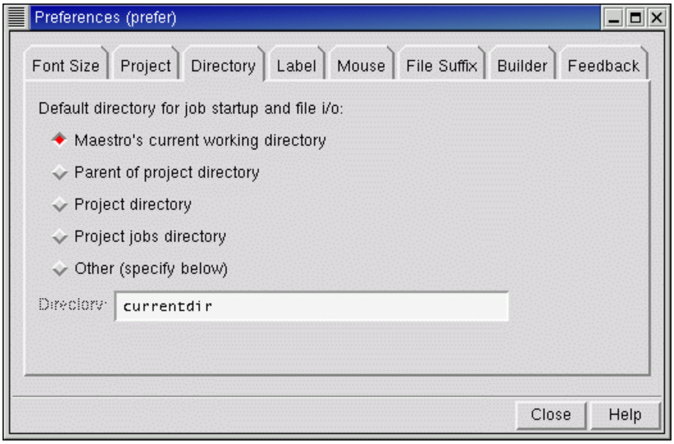


Figure 2.6. The* Directory *folder of the* Preferences *panel.

You can also set other preferences in the Preferences panel. See Section 12.2 of the *Maestro User Manual* for details.

2.9 Undoing an Operation

To undo a single operation, click the Undo button in the toolbar, choose Undo from the Edit menu, or press CTRL+Z. The word Undo in the menu is followed by text that describes the operation to undo. Not all operations can be undone: for example, global rotations and translations are not undoable operations. For such operations you can use the Save view and Restore view buttons in the toolbar, which save and restore a molecular orientation.

2.10 Running and Monitoring Jobs

Maestro has panels for each product for preparing and submitting jobs. To use these panels, choose the appropriate product and task from the Applications menu and its submenus. Set the appropriate options in the panel, then click Start to open the Start dialog box and set options for running the job. For a complete description of the Start dialog box associated with your computational program, see your product's User Manual. When you have finished setting the options, click Start to launch the job and open the Monitor panel.

The Monitor panel is the control panel for monitoring the progress of jobs and for pausing, resuming, or killing jobs. All jobs that belong to you can be displayed in the Monitor panel, whether or not they were started from Maestro. Subjobs are indented under their parent in the job list. The text pane shows output information from the monitored job, such as the contents of the log file. The Monitor panel opens automatically when you start a job. If it is not open, you can open it by choosing Monitor from the Applications menu in the Maestro main window.

While jobs are running, the Detach, Pause, Resume, Stop, Kill, and Update buttons are active. When there are no jobs currently running, only the Monitor and Delete buttons are active. These buttons act on the selected job. By default, only jobs started from the current project are shown. To show other jobs, deselect Show jobs from current project only.

When a monitored job ends, the results are incorporated into the project according to the settings used to launch the job. If a job that is not currently being monitored ends, you can select it in the Monitor panel and click Monitor to incorporate the results. Monitored jobs are incorporated only if they are part of the current project. You can monitor jobs that are not part of the current project, but their results are not incorporated. To add their results to a project, you must open the project and import the results.

Further information on job control, including configuring your site, monitoring jobs, running jobs, and job incorporation, can be found in the *Job Control Guide* and the *Installation Guide*.

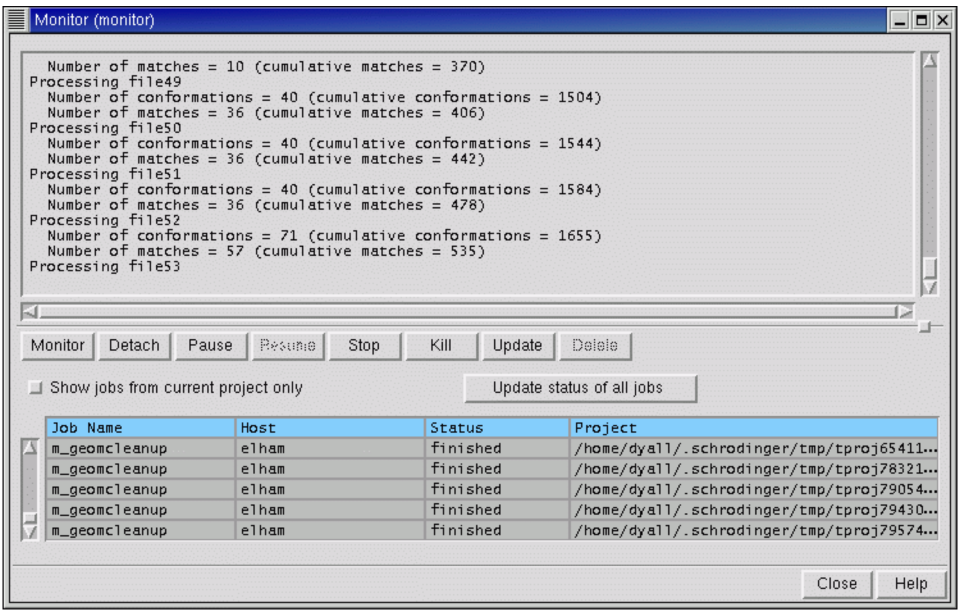


Figure 2.7. The Monitor ***panel.***

2.11 Getting Help

Maestro comes with automatic, context-sensitive help (Auto-Help), Balloon Help (tooltips), an online help facility, and a user manual. To get help, follow the steps below:

- Check the Auto-Help text box at the bottom of the main window. If help is available for the task you are performing, it is automatically displayed there. It describes what actions are needed to perform the task.
- If your question concerns a GUI element, such as a button or option, there may be Balloon Help for the item. Pause the cursor over the element. If the Balloon Help does not appear, check that Show Balloon Help is selected in the Help menu of the main window. If there is Balloon Help for the element, it appears within a few seconds.
- If you do not find the help you need using either of the steps above, click the Help button in the lower right corner of the appropriate panel. The Help panel is displayed with a relevant help topic.
- For help with a concept or action not associated with a panel, open the Help panel from the Help menu or press CTRL+H.

If you do not find the information you need in the Maestro help system, check the following sources:

- The *Maestro User Manual*
- The Frequently Asked Questions page on the Schrödinger Support Center.

You can also contact Schrödinger by e-mail or phone for help:

- E-mail: help@schrodinger.com
- Phone: (503) 299-1150

2.12 Ending a Maestro Session

To end a Maestro session, choose Quit from the Maestro menu. To save a log file with a record of all operations performed in the current session, click Quit, save log file in the Quit panel. This information can be useful to Schrödinger support staff when responding to any problem you report.

Building a Pharmacophore Model

This chapter is designed to help you become familiar with the Develop Pharmacophore Model workflow of Phase. This workflow involves the identification of common pharmacophore hypotheses from a set of active ligands. A common pharmacophore hypothesis is a spatial arrangement of chemical features common to two or more active ligands, which is proposed to explain the key interactions involved in ligand binding. Each hypothesis identified by Phase is scored according to how well the active ligands superimpose when they are aligned on the features associated with that hypothesis. A high-scoring hypothesis might be used to search a 3D database for new potentially active molecules, or it might be used to align a series of ligands in order to create a 3D QSAR model.

The structures and data used in this portion of the tutorial were taken from *J. Med. Chem.* **2003**, *46*, 716-726. The data set consists of 50 angiotensin AT_1 antagonists, divided into training and test sets of 25 ligands apiece. For convenience and clarity, ligand names have been assigned to indicate membership in the training set (train01, train02, ...) or the test set (test01, test02, ...). You will be developing pharmacophore and QSAR models from the training set, and applying them to the test set.

3.1 Preparing for the Exercises

To do the exercises, you must have access to an installed version of Maestro 7.5 and Phase 2.0. For installation instructions, see the *Installation Guide*. Before you start Maestro, you must create a working directory, and copy files from the Phase distribution into this directory.

To set up the working directory:

1. Change to a directory in which you have write permission.
2. Copy the archive file for the tutorial to this directory.

   `cp $SCHRODINGER/phase-v`*`version`*`/tutorial/phase_tutorial.tar.gz .`

3. Uncompress and extract the archive file:

   `gunzip -c phase_tutorial.tar.gz | tar xf -`

4. Change to the working directory for this exercise.

   `cd phase_tutorial/at1`

To prepare for running Phase and Maestro:

1. Set the `SCHRODINGER` environment variable to the directory in which Maestro and Phase are installed:

   csh/tcsh: `setenv SCHRODINGER` *installation_path*
   sh/bash/ksh: `export SCHRODINGER=`*installation_path*

2. Start Maestro with the command:

   `$SCHRODINGER/maestro &`

   The Maestro main window is displayed.

3.2 Importing Structures

1. In the Maestro main window, click the Import structures button on the toolbar.

   The Import panel is displayed.

2. Select the file `at1.mae`.
3. Click Import.
4. Click the Open/Close project table toolbar button.

   The Project Table panel is displayed. You should see a total of 50 molecules with various columns of experimental and predicted IC_{50} data for antagonism of the angiotensin AT_1 receptor. Of interest to us is the column labeled pIC50-Exp. These are the negative logarithms of the experimental IC_{50} values expressed in concentration units of mol/L. A value of 9.0 corresponds to an IC_{50} of 10^{-9} mol/L, or 1 nM.

3.3 Saving the Scratch Project

The project into which you imported the structures is a scratch project. Before proceeding, this project should be saved.

1. Click the Save As toolbar button.

The Save Project As project selector is displayed.

2. In the Project text box, type `at1_tutorial` at the end of the text.
3. Click Save.

3.4 Entering the Develop Pharmacophore Model Workflow

Maestro provides a wizard to guide you through the appropriate steps of the Develop Pharmacophore Model workflow, in the appropriate order. This workflow allows you to identify common pharmacophore hypotheses and create 3D QSAR models.

- Choose Develop Pharmacophore Model from the Phase submenu of the Applications menu in the main window.

  The Develop Pharmacophore Model panel is displayed.

At the bottom of the panel is a series of buttons labeled Prepare Ligands, Create Sites, and so on. These are the various steps in the Develop Pharmacophore Model workflow, and clicking on any enabled button takes you directly to that step. Currently, all buttons except Prepare Ligands are disabled. A button is enabled only if the prerequisite steps in the workflow have been successfully completed. The Step menu at the top of the panel can also be used to move around the workflow.

3.5 Adding Ligands

No structures have yet been added to the Phase workflow, so there are no entries in the Ligands table. The structures that are in the Project Table must now be added.

1. Click the From Project button, which is located near the top of the panel.

   The Add From Project dialog box is displayed.

   On the left side is a list of Project Table entry names, all of which are selected. If you wanted to add only a subset of the entries, you would simply make the desired selections in this dialog box. For now, ensure that all entries are selected.

   On the right side of the dialog box are the names of all the properties contained in the Project Table. If you intend to build QSAR models, this is where you should select the property to use as the experimental activity variable.
2. Select the pIC50-Exp property.
3. Click OK.

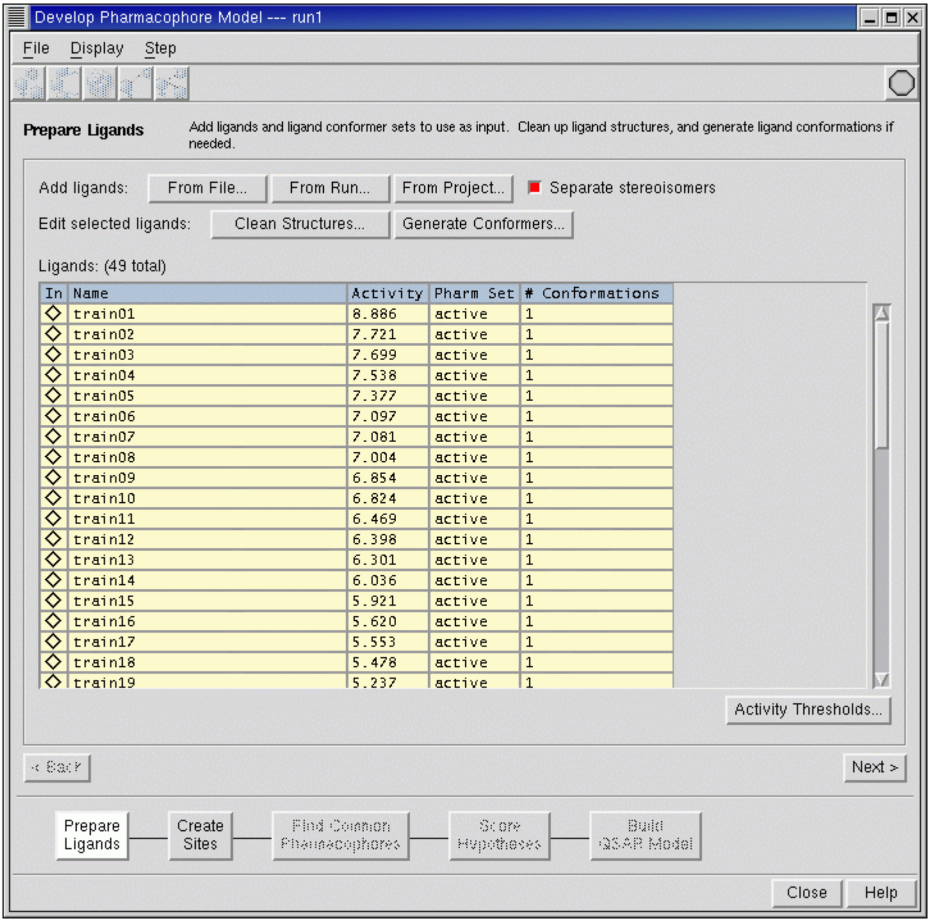


Figure 3.1. Ligands table in the* Prepare Ligands *step.

The Ligands table is now filled with *copies* of the structures from the Project Table. The fact that they are copies means that any subsequent changes made to the structures in the Project Table are not propagated to the corresponding structures in the Ligands table, and vice versa.

The Ligands table contains an In column, which allows you to view one or more ligands in the Maestro Workspace. The Name column holds the same information as the Title column of the Project Table, and the Activity column contains the pIC50-Exp data.

The Pharm Set column indicates whether a molecule is in the set of actives used to identify common pharmacophore hypotheses, or in the set of inactives used to eliminate nondiscriminatory hypotheses, or in neither set. If ligands of widely varying activity are present, you would normally want to use only the most active ones in the set of actives. The most active ligands are assumed to contain the strongest binding, most important, or greatest number of pharmacophore features that are involved in binding to the protein target. The set of actives should contain as much structural diversity as possible, so that the resulting pharmacophore models are applicable across different chemical families.

The # Conformations column indicates how many conformations are present for each ligand. When adding from the Project Table, you would normally see only a single conformation for each ligand. However, you may want to generate multi-conformer models outside Phase, export them to a file, then add them to Phase from that file.

Curiously, although 50 ligands were added from the Project Table, `test17` contains 2 conformations, and `test24` is missing from the Ligands table, even though it is present in the Project Table. What has happened here is that `test17` and `test24` were found to be chemically identical and were therefore merged into a single entry in the Ligands table. It turns out that these two structures are also identical in the publication from which they were taken, so you will proceed with a reduced data set of 49 ligands and the knowledge that the experimental activity value for `test17` might not be reliable.

3.6 Choosing the Active and Inactive Sets

As noted previously, only the most active compounds are normally considered when developing common pharmacophore hypotheses. Inactives can be used to eliminate hypotheses that do not provide a good explanation of activity on the basis of the pharmacophores alone. The active set determines the pool of pharmacophore models that are generated, and the initial scores that are assigned to them. The inactive set may be used subsequently to assign adjusted scores that reflect the degree to which the models distinguish actives from inactives. This is particularly useful if everything in the active set is built on a common scaffold, which can give rise to a number of spurious pharmacophore models that have nothing to do with ligand binding.

Accordingly, you will now set a threshold for actives of $IC_{50} \leq 50$ nM, which translates to $pIC_{50} \geq 7.3$, and a threshold for inactives of $pIC_{50} \leq 5.0$.

1. In the Develop Pharmacophore Model panel, click Activity Thresholds.

   The Activity Thresholds dialog box appears, which allows you to specify the activity thresholds.

2. In the Active if activity above text box, type `7.3`.

3. In the Inactive if activity below text box, type `5.0`.

4. Click OK.

The Pharm Set column now has active for each ligand whose pIC_{50} value is greater than 7.3, and inactive for each ligand whose pIC_{50} value is less than 5.0. The column is blank for ligands whose activity falls between these values.

5. Clear the Pharm Set column for any test set ligands that are assigned to the active or the inactive set.

   You can do this by selecting the assigned ligands and control-clicking in the column for one of ligands until the selected rows are blank in this column. The values cycle through active, inactive and blank when you click.

   The test set ligands should not be part of the pharm set because the pharm set is used to develop the pharmacophore model, and the test set is used to validate the model.

Five ligands should now be in the active set: `train01` through `train05`, and six ligands in the inactive set: `train20` through `train25`.

3.7 Cleaning Structures

The structures that were imported are 2D structures without explicit hydrogen atoms. To use them, hydrogens must be added and a single, low-energy, 3D conformation generated for each ligand. This task is performed by running LigPrep.

1. Ensure that all rows in the Ligands table are selected.

   You can do this by right-clicking in the table and choosing Select All from the shortcut menu.

2. Click Clean Structures.

   The Clean Structures dialog box is displayed. This dialog box contains a very basic subset of options that are passed to LigPrep. There are no chiral ligands in this data set, so you need not worry about the stereoisomer options.

3. Under Ionization, make sure Retain original states is selected.

4. Click Start.

5. When the green Start dialog box is displayed, click Start once again.

   The Clean Structures job has now been submitted. You can monitor the job by clicking on the spinning green octagon in the upper right corner of the Develop Pharmacophore Model panel. This job requires only about two minutes on a 2 GHz Pentium 4 processor.

   Once the job is finished and results are incorporated into the project, the flat structures in the Ligands table are replaced by single-conformer 3D models with the correct numbers of hydrogens attached to each atom.

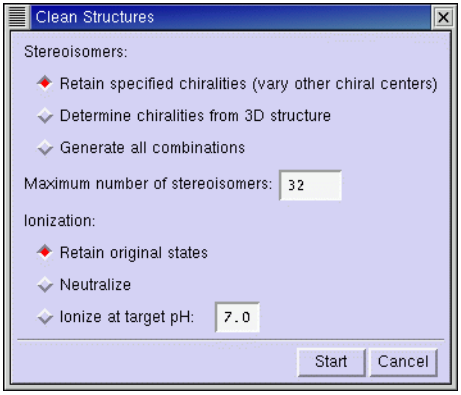


Figure 3.2. The Clean Structures ***dialog box.***

6. Verify that the structures have been cleaned by placing one or more ligands in the Workspace.

At this point, the interface allows you to advance to the Create Sites step, but there is only one conformation per ligand. While these are all low-energy models, there is no guarantee that they accurately reflect the geometries of the ligands when they bind to the receptor. In order to find meaningful common pharmacophore hypotheses when the geometry of the ligands in the active site is not known, a range of conformations that are accessible to each ligand must be sampled.

3.8 Creating Conformations

The next task is to generate sets of conformers for each ligand. You can generate conformers for each ligand by one of two methods: a ligand torsion search, which involves systematic sampling around rotatable bonds (as in Glide), or a mixed Monte-Carlo multiple minimization/low-mode search (MCMM/LMOD). Both methods can be followed by MacroModel minimization and filtering.

1. Ensure that all rows in the Ligands table are selected.

   Conformers are only generated for the selected ligands.

2. Click Generate Conformers.

   The Generate Conformers dialog box opens.

3. For Current conformations, select Discard.

4. For Sampling, select Rapid.

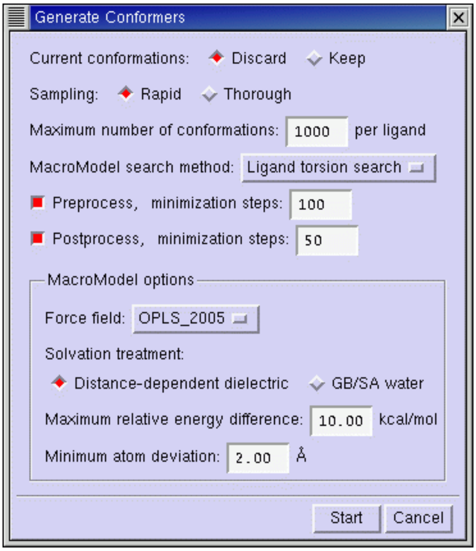


Figure 3.3. The Generate Conformers ***dialog box.***

5. Set the maximum number of conformations to 50.
6. Select Ligand torsion search from the MacroModel search method option menu.
7. Ensure that Preprocess and Postprocess are both selected, and that the values in the text boxes are 100 and 50, respectively.
8. Ensure that the force field is OPLS_2005.
9. Under Solvent treatment, select Distance-dependent dielectric.
10. Set the maximum relative energy difference to 10.0 kcal/mol.
11. Set the minimum atom deviation to 1.0 .
12. Click Start in this dialog box and again in the green Start dialog box.

This job requires about 30 minutes on a 2 GHz Pentium 4 processor. After the job finishes and results are incorporated, the Ligands table indicates roughly 30-40 conformations per ligand. Only the first conformation is viewable in the Workspace, but later in the workflow it will be possible to view other conformations. You are now ready to proceed to the Create Sites step.

Note that for high precision work, you might want to select the Thorough search mode and specify a larger number of conformations. A good rule of thumb is 50-100 conformations for each rotatable bond.

3.9 Proceeding to Create Sites

After creating conformations for each ligand, the next task is to identify and store the pharmacophore sites for each conformation.

- Click Create Sites in the Guide, or click Next.

There may be a delay of a few seconds before the panel is updated to the next step, while Phase does a fast mapping of pharmacophore features to the first conformation of each ligand. This mapping allows you to preview features before generating them for all conformations.

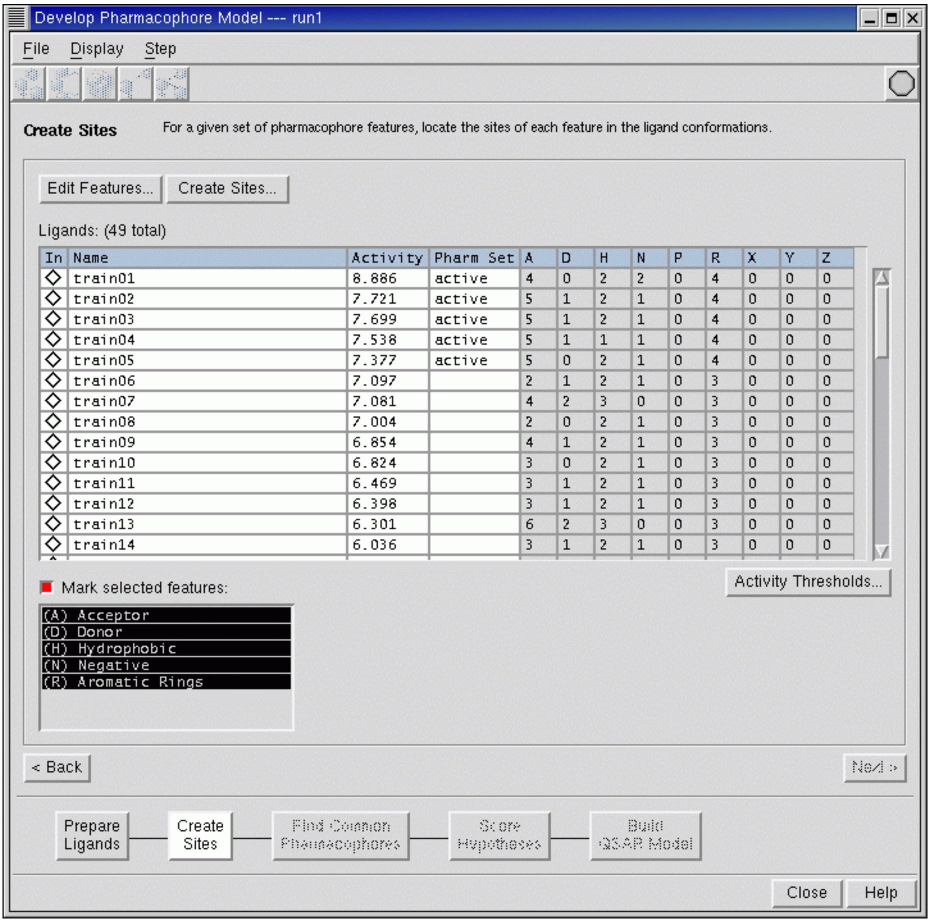


Figure 3.4. The Create Sites ***step.***

Once the panel is updated, the Ligands table contain columns (A, D, N, etc.) showing the counts of each type of pharmacophore feature in each ligand. By default, Mark selected features is selected, and all features in the list below it are selected.

3.10 Examining Feature Mappings

To fully understand how pharmacophore models are developed, it is important to understand how the various pharmacophore features are mapped onto the ligand structures.

Note: This exercise is optional, and does not affect the final pharmacophore model.

1. Click the Clear Workspace button on the main toolbar.

2. Select (A) Acceptor in the features list under Mark selected features.
3. Place the third ligand (`train03`) in the Workspace by clicking the In check box in the third row of the Ligands table.

   You should see various pink transparent spheres with arrows protruding from them. The spheres are centered on the hydrogen bond acceptor features of the ligand, and the arrows indicate the axes along which ideal hydrogen bonds would be formed.

4. Select (D) Donor in the features list to display hydrogen-bond donors.

   A blue sphere appears on the hydroxyl hydrogen of this ligand. Once again, the arrow indicates the direction of the ideal hydrogen bond.

5. Select (H) Hydrophobic in the features list to display hydrophobic features.

   A green sphere appears in the middle of the n-butyl chain attached to the imidazole ring. Since there is no directionality to the hydrophobic feature, it is represented as a sphere without an arrow.

6. Select (N) Negative in the features list to display negatively-charged features.

   A red sphere appears in the center of the tetrazole ring. While there is no ionic charge in this ligand structure, tetrazole is known to be acidic, and so it is perceived as a negative ionizable feature, much like a carboxylic acid.

7. Select (R) Aromatic Rings in the features list.

   Orange toroids illustrate the locations of aromatic rings.

8. Select (A) Acceptor in the features list again.

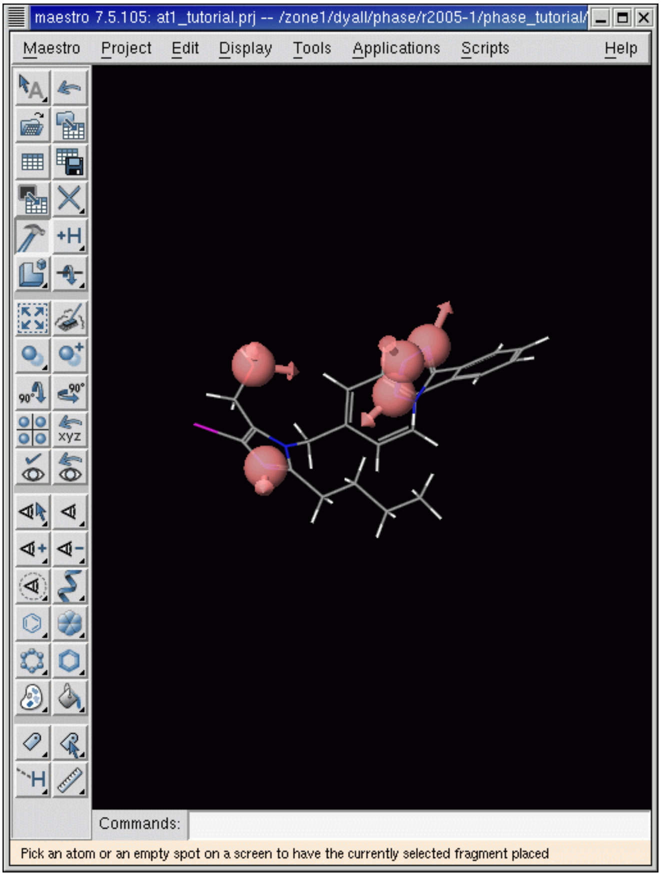


Figure 3.5. Acceptor features for ligand `train03`.

3.11 Examining and Modifying Feature Definitions

The pharmacophore features that were visualized in the previous section are mapped to the structures using a set of topological feature definitions, which you will examine here.

1. Click Edit Features.

   The Edit Features dialog box is displayed. This dialog box contains SMARTS patterns and other information that control the application of pharmacophore feature definitions. By default, definitions for the Acceptor (A) feature type are displayed; other feature types may be viewed by selecting from the Feature option menu. These are the built-in definitions provided with every installation of Phase, and they cannot be modified or deleted. They can, however, be *ignored*, as you will see.

Definitions near the top of the table have higher precedence than those closer to the bottom of the table. So, for example, if the first pattern maps a particular nitrogen in the ligand as an acceptor, that same nitrogen will not be mapped as an acceptor by any subsequent pattern. The vertical positions of the built-in features may not be changed, but user-defined patterns (discussed in the next section) may be moved using the arrowhead buttons below the Pattern list table.

It should be noted that the order of precedence does not apply to *excluded* patterns, i.e., patterns for which a check mark appears in the Exclude column. Excluded patterns, which will be discussed shortly, are processed before all other patterns.

To understand how the definitions are applied, consider the seventh entry in the table, for which the SMARTS pattern is `[n;X2]([a])([a])`. This 3-atom pattern is matched by an aromatic nitrogen and the two aromatic atoms to which it bonds.

2. Click the Mark box in the seventh row.

   The atoms and bonds in the matching substructures are now marked in pink on the Workspace ligand (`train03`). The pink acceptor spheres, which should also be visible, clarify that the pattern is matched three times in the tetrazole ring.

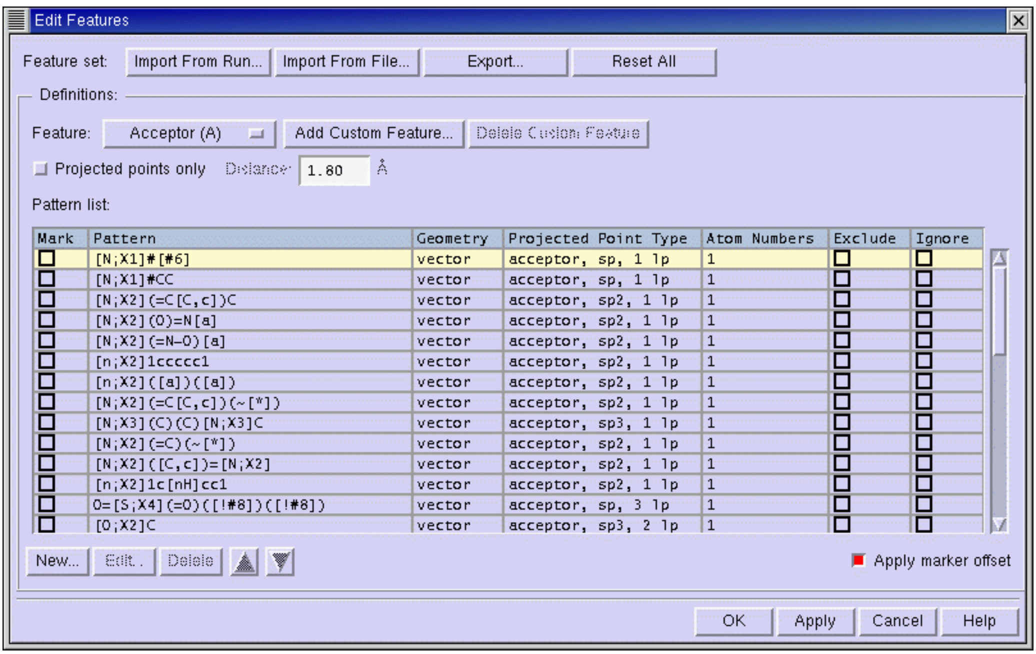


Figure 3.6. The Edit Features ***dialog box.***

The Geometry for this feature definition is vector, meaning that the feature is located on a single atom and has one or more directions associated with it—the directions of the possible hydrogen bonds. Hydrogen bond acceptors and donors are vector features, and aromatic rings are also vector features because the orientation of the ring is important. The other available geometries are point and group. These have no directionality and the associated features are located on either a single atom (point) or at the centroid of a group of atoms (group). Atom Numbers are referenced to the SMARTS pattern and are used in conjunction with Geometry to define the location of the feature. Projected Point Type designates the configuration of idealized hydrogen bonds for determination of vector orientations in acceptor and donor features.

Some feature definitions are checked as Exclude. Exclusion performs a logical NOT operation for matches to these patterns. To clarify what this means, the following steps show how Exclude definitions are applied.

3. Clear the Mark box for the previous SMARTS pattern `[n;X2]([a])([a])`.

   The substructure markers disappear from the Workspace ligand.

4. Remove `train03` from the Workspace and replace it with `train01` (by clicking the In check box for `train01` in the Ligands table).

   The Workspace displays the ligand `train01` and its acceptor features.

5. In the Edit Features dialog box, click the Mark box next to the SMARTS pattern `[O;X1]=[C,c]`, which is about half-way down the Pattern list table.

   The C=O bond in the carboxylic acid in `train01` is now marked. However, there is no pink sphere on the oxygen atom, so it is evidently not being perceived as an acceptor feature. This is because the built-in definitions assume that the user will *not* want to treat this as an acceptor due to the fact that the –COOH group is likely to be ionized. Accordingly, there is an *excluded* pattern further down in the table that matches this type of oxygen and excludes it from being flagged as an acceptor.

6. Clear the Mark box next to the SMARTS pattern `[O;X1]=[C,c]`.

   The C=O bond in `train01` is no longer marked.

7. Scroll down the table until you find the SMARTS pattern `O=C[O-,OH]`.

   The Exclude box is checked, which means this is an excluded (NOT) pattern.

8. Click the Mark box next to this pattern.

   The O=C–O moiety is now marked on `train01`. Since the Geometry of this pattern is point (that is, the feature is located on a single atom), it excludes only features that are also located on a single atom, that is, features with point and vector geometries. There is no entry in the Atom Numbers column, so the location of the feature defaults to the first atom in the pattern, which is the double bonded oxygen. So this exclude pattern matches all O=C–O moieties, and prevents the double bonded oxygen in -COOH from being flagged as an acceptor by any other pattern. As noted previously, excluded patterns are actually processed first, so the double bonded oxygen would be flagged for exclusion before any of the regular acceptor patterns were applied.

For the AT_1 data set, almost all of the ligands have an acceptor feature coming off the imidazole ring at the same position as the –COOH group in `train01`, indicating that there is probably an important hydrogen-bonding interaction going on. Moreover, `train01` is the most active ligand in the data set and it differs from the second most active ligand (`train02`) only by replacement of a $–CH_2OH$ group with –COOH. It would appear, then, that –COOH strengthens this hydrogen bonding interaction. Therefore, when looking for common pharmacophore hypotheses, it may be wise to allow at least one of the oxygens in this –COOH to act as an acceptor. Accordingly, you will *ignore* this exclude definition, so that it is not applied.

9. In the Ignore column, click the box for the `O=C[O-,OH]` pattern.

   The ignore operation is equivalent to removing the definition from the table, although you can always reinstate it by clearing the Ignore check box.

10. Click Apply at the bottom of the panel to update the definitions.

    A delay occurs as the new feature definitions are applied to the first conformation of each ligand. When this process is complete, a pink acceptor sphere appears on the oxygen in the C=O moiety. The excluded pattern definition `O=C[O-,OH]` is now being ignored, so that –COOH groups will be perceived as containing an acceptor feature.

It is important to note that the built-in definitions used to identify hydrophobic features (H) and aromatic rings (R) are not based on SMARTS patterns. Rather, special algorithms are applied to detect these features automatically, and more efficiently, than would be possible using SMARTS patterns. However, it is possible to add new SMARTS-based definitions to any of the feature types, and to add new features with their own definitions. These possibilities are explored in the next two optional exercises.

3.12 Adding New Feature Definitions

The built-in feature definitions are reasonable, but there may be times when you need to add your own definitions to identify features that aren’t accounted for by the built-in set. In this exercise you will define a new type of acceptor feature. The feature you will add is an aromatic ring, which can function as a weak acceptor.

Note: This exercise is optional, and does not affect the final pharmacophore model.

1. Select (A) Acceptor in the features list.

   The acceptor features are displayed in the Workspace.

2. In the Edit Features dialog box, select Acceptor (A) from the Feature option menu.
3. Click the first row in the Pattern list table.
4. Click New.

   The New Pattern dialog box is displayed.

5. In the SMARTS pattern text box, type `c1ccccc1`.
6. Choose Group from the Geometry option menu.

   The geometry in this case is defined by a group of atoms that together contribute an acceptor rather than a single atom.

7. Choose Aromatic ring from the Projected point type option menu.

   Most acceptors have a directionality that can be precisely defined by considering the orientation of the acceptor lone pairs with respect to some plane in which the atom lies. In this case the direction is perpendicular to the ring.

8. Click OK.

   A new acceptor pattern appears at the top of the Pattern list table. In this position in the table, this pattern is matched before all others. You could move this new pattern vertically in the table using the arrowhead buttons, but doing so ultimately has no effect on the feature perception since there is no other similar functional group.

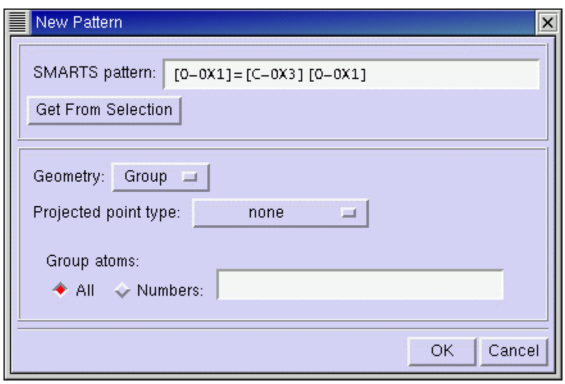


Figure 3.7. The New Pattern ***dialog box.***

9. Click the Mark box for this new definition.

   The two occurrences of aromatic rings are marked in the Workspace.

10. Click Apply at the bottom of the panel to update the definitions.

    After a short delay, a pink sphere appears on each of the rings, indicating that they will be treated as an acceptor. This was for demonstration purposes only, so we should remove the definition before proceeding.

11. Select the newly created pattern in the Pattern list table.

12. Click the Delete button below the table to remove the definition permanently.

13. Click Apply at the bottom of the panel.

    The rings are no longer perceived as acceptors. You could have accomplished the same thing by simply *ignoring* the new definition, without permanently deleting it.

3.13 Adding New Feature Types

It is also possible to create an entirely new type of feature (X, Y or Z) beyond the set of built-in feature types (A, D, H, N, P, R). In this exercise, you will create a feature that identifies aromatic rings as weak H-bond acceptors. In practice, you might want to create feature types for weak acceptors and strong acceptors so that the two can be separated in a pharmacophore model. Here, to define the SMARTS pattern, you will use Workspace selection.

Note: This exercise is optional, and does not affect the final pharmacophore model.

1. In the Edit Features panel, click Add Custom.

   The Add Custom Feature dialog box is displayed.

2. Change the name from `Custom1` to `WeakAcceptor`.

3. Click OK.

   The feature WeakAcceptor (X) has been created and an empty Pattern list appears.

4. In the Workspace, select the six carbon atoms in an aromatic ring.

   You can do this using shift-click or by dragging over the atoms.

5. Click the New button below the Pattern list table.

   The New Pattern dialog box is displayed.

6. Click Get From Selection.

   A SMARTS pattern for the selected atoms appears in the SMARTS pattern text box.

7. Ensure that Geometry is Group, Projected point type is aromatic ring and Group atoms is All.
8. Click OK.

   The new pattern appears in the Pattern list table.
9. Check the Mark box next to the new pattern.

   The aromatic rings are marked in the Workspace.
10. Click Apply at the bottom of the Edit Features dialog box.
11. In the Develop Pharmacophore Model panel, click (X) WeakAcceptor to display occurrences of this new feature.

    After a brief delay, a turquoise sphere appears at the centroid of each ring, indicating that the new X feature type has been perceived.

This example was for demonstration purposes only, so you should remove the custom feature X before proceeding.

12. In the Edit Features dialog box, ensure that the custom WeakAcceptor (X) definitions are displayed.
13. Click Delete Custom Feature.

    The custom WeakAcceptor (X) definitions are removed and the Acceptor (A) definitions are displayed.
14. Click OK.

    The Edit Features panel is closed and the updated definitions are stored.
15. Verify that the custom X feature is no longer visible

You are now ready to apply the feature definitions to create pharmacophore sites for all conformations.

3.14 Creating Pharmacophore Sites

All of the operations done thus far applied the feature definitions only to the first conformation of each ligand. Before you can proceed to the next step in the workflow, you must apply the feature definitions to all conformations. Pharmacophore sites are created for all entries in the Ligands table, regardless of whether any rows are selected.

1. Click the Create Sites button near the top of the panel (not the one in the Guide).
2. Click Start in the green Start dialog box.

This job requires less than a minute on a 2 GHz Pentium 4 processor. Incorporation of results does not add any new information to the Ligands table, but the Find Common Pharmacophores button at the bottom of the panel is now active, and you can proceed to the next step in the workflow.

3.15 Proceeding to Find Common Pharmacophores

In this part of the workflow, pharmacophores from all conformations of the active ligands are examined, and those pharmacophores that contain identical sets of features with very similar spatial arrangements are grouped together. If a given group is found to contain at least one pharmacophore from each active ligand, then this group gives rise to a *common pharmacophore*. Any single pharmacophore in the group could ultimately become a common pharmacophore *hypothesis*.

- Click either Find Common Pharmacophores or Next.

  The Find Common Pharmacophores step is displayed.

By default, Phase looks for common *5-point* pharmacophores, that is, pharmacophores containing 5 sites. However, the number of sites can be set to any value between 3 and 7. Be aware, though, that as the number of sites is increased, the likelihood of finding common pharmacophores is decreased.

Also by default, Phase looks for pharmacophores that are common to all active ligands. However, Phase allows you to relax this criterion so that a common pharmacophore need only match a subset of the active ligands. This is often a necessity when the actives are highly diverse. So in general, a common pharmacophore must match a *minimum required number of actives*, where that minimum number is set by the user. The present set of five actives is not very diverse, so leave the value in the Must Match at least option at 5.

In the Feature frequencies table, there is a column labeled Available. This is the upper limit on the number of times each feature could possibly be found in any common pharmacophore. For example, the table indicates that no more than one negative feature (N) can appear in any common pharmacophore. This value is arrived at by identifying the active ligand with the smallest number of negative features. If the minimum required number of actives is reduced, then Available values may increase, according to whether or not the most feature-deficient ligands could be excluded from the subset.

The Minimum and Maximum columns in the Feature frequencies table contain user-settable limits on the number of times a feature is *allowed* to appear in a common pharmacophore. For example, if we wanted common pharmacophores to contain at least one, but no more than three acceptors (A), we would change the Minimum and Maximum limits on this feature to 1 and 3.

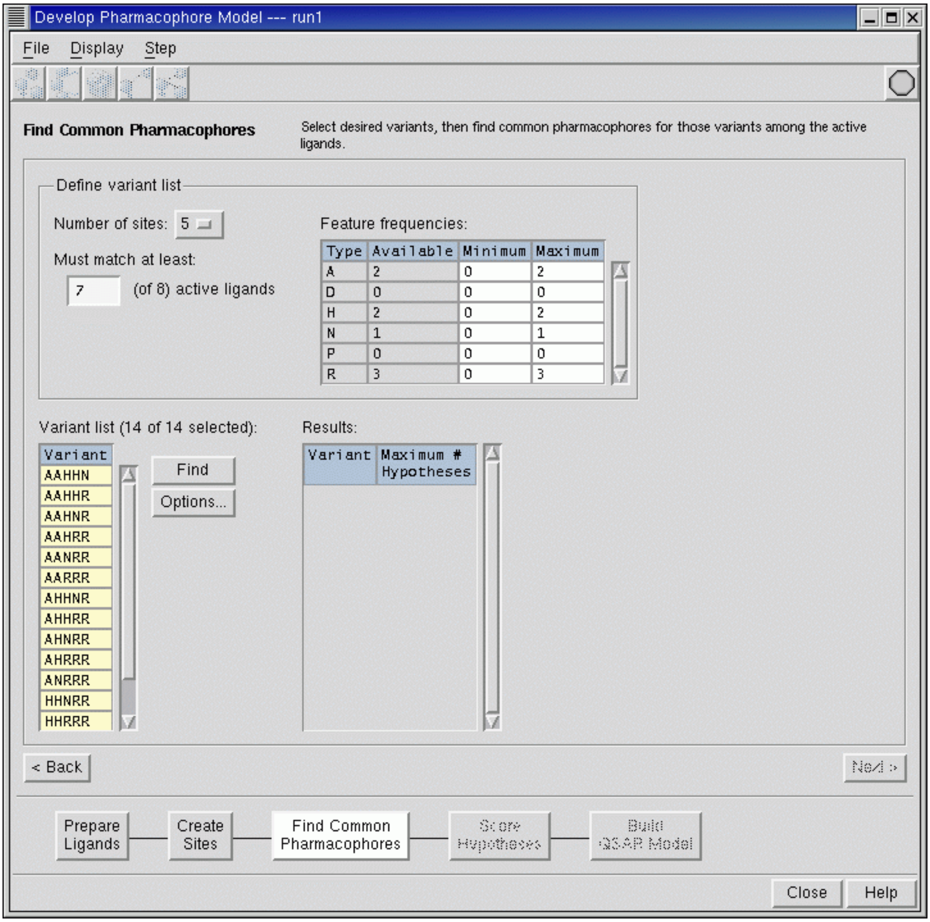


Figure 3.8. The Find Common Pharmacophores ***step.***

A list of 19 *variants* also appears on the panel, reflecting the 19 possible combinations of features that could give rise to common pharmacophores. For example, AAARR indicates that the potential exists for common pharmacophores containing three acceptors and two aromatic rings. The feature combinations you see here are determined entirely by the Minimum and Maximum limits in the Feature frequencies table.

3.16 Changing the Number of Sites in Common Pharmacophores

This exercise examines how the features frequencies and variants change with the number of sites in the common pharmacophores we are searching for.

1. Choose 4 from the Number of sites option menu.

   The only change in the Feature frequencies table is that the Maximum allowed frequency of acceptors has been reduced from 5 to 4. This merely reflects the fact that the common pharmacophores we are looking for cannot contain more than 4 sites. The number of variants has dropped from 19 to 16, and of course there are only 4 features in each variant.

2. Change the number of sites to 6.

   The Feature frequencies table is identical to that observed when the number of sites was 5, and there are once again 19 variants.

3. Change the number of sites to 7.

   The Feature frequencies table is unchanged, but the number of variants drops to 16. Combinatorics dictates that there will be a peak in the number of variants as the number of sites is varied, and we see that the peak occurs at 5 and 6 sites. You will be searching for common 6-point pharmacophores.

4. Change the number of sites back to 6.

3.17 Changing the Allowed Feature Frequencies

The Feature frequencies table and the variant list indicate that we can search for common pharmacophores with up to five acceptors. However, it is very unlikely that a ligand would bind to the receptor through five, or even four, hydrogen bond acceptor interactions. Here, some chemical intuition is needed to filter out results that are scientifically unsound or unlikely.

1. In the Feature frequencies table, change the Minimum and Maximum number of acceptors to 1 and 3, respectively.

   The number of variants drops from 19 to 11 and each variant contains between 1 and 3 acceptor features.

We know that each active ligand contains a tetrazole ring and therefore at least one negative feature. The tetrazole has no doubt been put there for a good reason, so it is pretty safe to assume that the hypothesis should contain a negative feature.

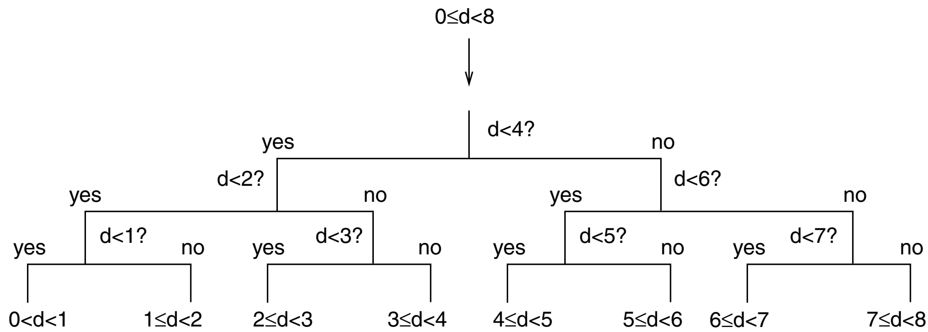


Figure 3.9. Binary decision tree.

2. In the Feature frequencies table, change the Minimum number of negative features from 0 to 1.

   The number of variants drops from 11 to 6, and each variant contains a negative feature.

The Feature frequencies table also indicates that there could be as many as 4 aromatic rings. However, one of these rings is the tetrazole, which we presume will be acting as a negative feature, so it is not necessary to consider so many aromatic rings.

3. In the Feature frequencies table, change the Maximum number of aromatic rings from 4 to 3.

   The number of variants drops from 6 to 5, and no variant contains more than 3 aromatic rings.

3.18 Examining Options for Finding Common Pharmacophores

To understand the options for finding common pharmacophores, we must first examine how the underlying algorithm works. Common pharmacophores are identified using a tree-based partitioning technique that groups together similar pharmacophores according to their *intersite distances*, i.e., the distances between pairs of sites in the pharmacophore. Accordingly, each k-point pharmacophore is represented by a vector of n distances, where $n = k \cdot (k-1)/2$. Each intersite distance d is filtered through a binary decision tree, such as in Figure 3.9.

Note: This exercise is optional, and does not affect the final pharmacophore model.

The tree in Figure 3.9 has a depth of three and partitions distances (in angstroms) on the interval $0 < d \leq 8$ into bins that are 1 Å wide. If each of the n distances in a pharmacophore is filtered in this manner, an n-dimensional partitioning of the pharmacophore is created.

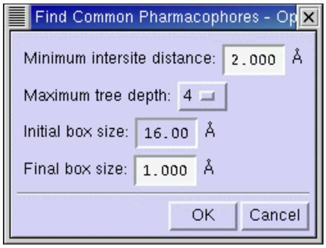


Figure 3.10. The Find Common Pharmacophores - Options ***dialog box.***

This representation is referred to as an *n*-dimensional box, where the sides of the box are equal to the bin width. Thus a pharmacophore is mapped, according to its intersite distances, into a box of finite size. All pharmacophores that are mapped into the same box are considered to be similar enough to facilitate identification of a common pharmacophore. So if each of a minimum required number of actives contributes at least one pharmacophore to a particular box, then that box represents a common pharmacophore. Such boxes are said to *survive* the partitioning procedure, while all others are eliminated.

Armed with this information, you may now proceed to examine the options for partitioning pharmacophores into boxes.

1. Click Options.

   The Find Common Pharmacophores - Options dialog box is displayed. Most of these options can be directly linked to the example tree in Figure 3.9. Initial box size refers to the width of the initial distance interval. By default this is 16 Å ,indicating that intersite distances on the interval $0 < d \leq 16$ will be processed. Final box size refers to the size of the *n*-dimensional boxes into which the intersite distances are ultimately mapped. The default is 1 Å .Note that Initial box size is not actually set by the user, but rather computed from Maximum tree depth and Final box size:

   Initial box size = (Final box size) $\times\ 2^{(\text{Maximum tree depth})}$

2. To verify the relationship among Initial box size, Final box size and Maximum tree depth, change Maximum tree depth to 5.

   Initial box size increases to 32.0.

   The only remaining option is Minimum intersite distance. This parameter determines whether or not a pharmacophore will be rejected because a pair of features is too close together. For example, suppose a ligand contains a –CH_2OH group. Using the built-in feature definitions, an acceptor feature would be placed on the oxygen and a donor feature would be placed on the hydrogen. While it is theoretically possible for an oxygen atom to act as both donor and acceptor (e.g., water), it is probably unlikely in a ligand-receptor interaction. Accordingly, we would typically want to ignore pharmacophores that contain both the acceptor and donor features associated with a single oxygen atom. Since these two features are separated by only about 1 Å, a Minimum intersite distance of 2.0 Å would cause the entire pharmacophore to be rejected before it is even run through the partitioning tree. The default value of 2.0 Å is recommended, as are all other defaults in this dialog box.

3. Click Cancel to close this dialog box and keep all default settings.

You are now ready to search for common pharmacophores.

3.19 Finding Common Pharmacophores

A total of five variants are examined to identify 6-point pharmacophores that are common to all five of the active ligands.

1. Ensure that all five variants are selected, then click Find.
2. Click Start in the green dialog box.

This job requires about two minutes on a 2 GHz Pentium 4 processor. After the job incorporates, the Results table shows that only two of the five variants yielded common pharmacophores. Maximum # Hypotheses is the number of different 1-Å boxes that survived the partitioning procedure, and hence the number of distinct common pharmacophores that were identified for each variant. Recall that a given box contains one or more pharmacophores from each of the minimum required number of active ligands. Exactly one pharmacophore from each box will be selected as a potential hypothesis, and this selection takes place in the Score Hypotheses step.

3.20 Proceeding to Score Hypotheses

In the next step, common pharmacophores are examined, and a scoring procedure is applied to identify the pharmacophore from each box that yields the best alignment of the active ligands. This pharmacophore provides a hypothesis to explain how the active ligands bind to the receptor. There will of course be many hypotheses, because there are many boxes. The scoring procedure provides a ranking of the different hypotheses, allowing you to make rational choices about which hypotheses are most appropriate for further investigation.

- Click either Score Hypotheses or Next.

  The Score Hypotheses step is displayed. The Hypotheses table is empty because scoring has not yet been done.

3.21 Scoring Hypotheses

In this exercise, you will examine settings that control how hypotheses are selected from each surviving box and how they are ranked with respect to one another, and then run scoring jobs to examine the effect of including various terms in the score.

1. Click Score Actives.

   The Score Actives dialog box is displayed.

To understand the options in this dialog box, you must first understand the scoring process. A surviving box contains a set of very similar pharmacophores culled from conformations of a minimum number of active ligands, and certain of these ligands may contribute more than one pharmacophore to a box. Each pharmacophore and its associated ligand are treated temporarily as a *reference* in order to assign a score. This means the other *non-reference* pharmacophores in the box are aligned, one-by-one, to the reference pharmacophore, using a standard least-squares procedure applied to the corresponding pairs of site points.

At this stage, the quality of each alignment is measured using up to three terms: (1) the root-mean-squared deviation (RMSD) in the site point positions; (2) the average cosine of the angles formed by corresponding pairs of vector features (acceptors, donors and aromatic rings); and (3) a volume overlay term based on van der Waals models of the non-hydrogen atoms in each pair of structures.

$$S_{\mathrm{vol}}(i) = V_{\mathrm{common}}(i)/V_{\mathrm{total}}(i)$$

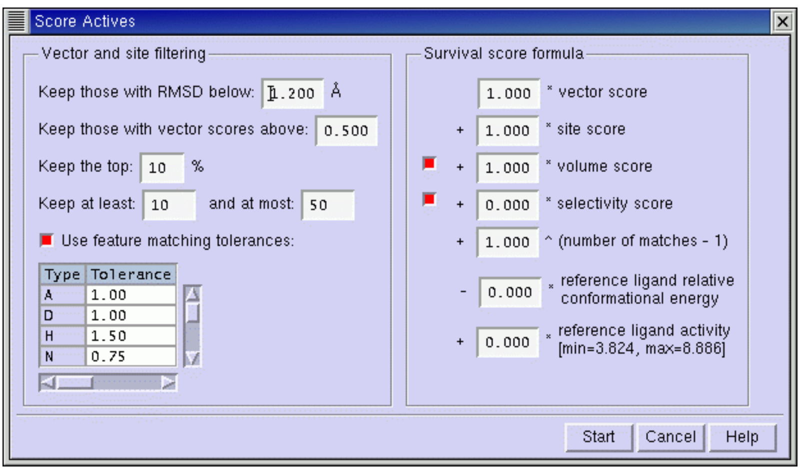


Figure 3.11. The Score Actives ***dialog box.***

$V_{common}(i)$ is the common or overlapping volume between ligand *i* and the reference ligand, while $V_{total}(i)$ is the total volume occupied by both ligands. Note that the volume term is computed only if the option to score by volume is selected. These two or three terms are combined with separate weights to yield a combined alignment score for each non-reference pharmacophore that has been aligned to the reference. If a non-reference ligand contributes more than one pharmacophore to the box, the pharmacophore yielding the best alignment to the reference is selected. The overall multi-ligand alignment score for a given reference pharmacophore is the average score from the best set of individual alignments.

In principle, a reference pharmacophore could yield a good average score, even though it contains one or two very poor individual alignments. For this reason, user-adjustable cutoffs are applied to the RMSD values and vector cosines of each individual alignment. Any reference pharmacophore that violates a cutoff in any individual alignment is eliminated.

After all pharmacophores in a box have been treated as a reference, the one yielding the highest multi-ligand alignment score is selected as the hypothesis for that box. The ligand that contributes the reference pharmacophore is referred to as the *reference ligand* for that hypothesis. Note that the non-reference information is carried along with each hypothesis so that additional scoring may be performed using the optimal multi-ligand alignment.

Once hypotheses have been identified across all boxes, you may want to eliminate some of the lower scoring ones. Accordingly, a percentage cutoff may be applied to the overall alignment score so that, for example, only hypotheses in the top 10% are retained. You may also request that some minimum number of hypotheses be retained, just in case the percentage filter yields a very small number.

Once this stage of scoring is completed, you can further refine the ranking of the hypotheses by adding to the scoring function terms for selectivity, reference ligand relative energy, and reference ligand activity. You might also want to weight hypotheses that match more actives.

Selectivity is an empirical estimate of the *rarity* of a hypothesis, i.e., what fraction of molecules are likely to match the hypothesis, regardless of their activity toward the receptor. Selectivity is defined on a logarithmic scale, so a value of 2 means that $1/10^2$ molecules would be expected to match the hypothesis. Higher selectivity is desirable because it indicates that the hypothesis is more likely to be unique to the active ligands. Selectivity is only a rough estimate of the rarity, so you should be careful not to place too much emphasis on it in the overall ranking of hypotheses. As with the other types of scores, selectivity can be added to the overall score with its own adjustable weight.

You may also wish to assign higher scores to hypotheses that match a greater number of active ligands. This is relevant when the required minimum number of actives is smaller than the total number of actives. The reward comes in the form of w^m, where *w* is adjustable (1.0 by default) and m is the number of actives that match the hypothesis. If w is increased above 1.0, care must be taken not to make it too large, or it may completely dominate the scoring function.

Logically, the best hypothesis should have as a reference ligand the ligand that has the highest activity. You can weight the score towards this goal by including the activity in the scoring.

If the conformation of the reference ligand is high in energy with respect to the lowest conformation, this can indicate a poor hypothesis, because the internal strain of the ligand must be overcome when the ligand binds. You can include a penalty term by subtracting off a multiple of the reference ligand relative energy.

By default, none of these optional scores contribute to the ranking of the hypotheses. Although selectivity score is selected as part of the scoring function, its weight is zero. However, all individual components of the score are reported, so you can always consider selectivity, if you so desire, when making choices among hypotheses that have very similar overall scores.

2. Check that the weights of the selectivity score and the number of matches are 0.0 and 1.0.
3. Click Start, and click Start again in the green Start dialog box.

   This job requires less than one minute on a 2 GHz Pentium 4 processor. Once the job is incorporated, the Hypotheses table contains information about the highest scoring hypotheses from each variant.

   The highest-scoring hypotheses are `AAHNRR.68`, `AHNRRR.197`. However, the range of scores among the reported hypotheses is only 0.5, so none of the hypotheses is particularly poor.

Next, you will check whether any of these hypotheses can be eliminated based on its match to the inactives.

4. Click Score Inactives.

   The Score Inactives dialog box opens. Leave the inactives weight at its default value of 1.0.

5. Click Start, and click Start again in the green Start dialog box.

   The job takes less than a minute. When it finishes, the Survival-inactive column of the Hypotheses table is populated with values.

   The survival score is reduced fairly evenly for all the hypotheses. A reduction of one unit is expected, because the scoring function for the inactives is the same as for the actives, and it includes a contribution of 1.0 for the number of matches. The remainder of the reduction is fairly uniform and fairly small, indicating that the inactives do not match the hypotheses well enough to eliminate any of them. On this basis, we can be fairly confident that the hypotheses are reasonable explanations of the activity.

Finally, you will adjust the scoring to take account of the activity. Of the five actives used for model development, one has a significantly higher activity than the others. Adding an activity reward could change the ranking of the hypotheses that have this ligand as the reference.

6. Click Rescore.

   The Rescore Hypotheses dialog box opens.

7. Enter `0.3` in the reference ligand activity text box.

8. Click OK.

   The results are returned almost immediately. Scoring the actives and the inactives requires alignment of the ligands, which takes a little time. For rescoring the alignment is already done.

9. Click twice on the Post-hoc column heading, to sort the hypotheses in descending order by this score.

   Now, `AHNRRR.53` and `AAHNRR.55` are at the top, followed by the set of hypotheses with the top survival scores. In the next exercise, you will examine some of these hypotheses and the ligand alignments.

3.22 Viewing Hypotheses and Ligand Alignments

In this exercise you will examine the nature of the top-scoring hypotheses and the quality of the associated ligand alignments.

1. Sort the Hypotheses table by survival score, by clicking twice on the Survival column heading.

   The top scoring hypothesis is identified as `AAHNRR.68`. The suffix 68 is merely the index of the box from which the hypothesis came. The second highest scoring hypothesis, `AHNRRR.197`, is associated with a different variant, but it has scores that are nearly identical to those of `AAHNRR.68`. As we shall see, this is due to the fact that both hypotheses align the active ligands in very similar ways.

2. Select the first row in the Hypotheses table.

   The features of the hypothesis are displayed in the Workspace and the Alignments table is filled with a record for each ligand. The records for the ligands not used for the model (the *non-model* ligands) are dark gray, indicating that no alignment was done for these ligands. The records provide information on the conformation whose pharmacophores yielded the best multi-ligand alignment to the hypothesis when it was selected as the reference from its box. Fitness measures the quality of each alignment using a weighted sum of alignment and volume scores, just as in the total scoring function. Note that the second row of the Alignments table is blue, indicating that `train02` is the reference ligand, i.e., the ligand from which the hypothesis came. Its fitness score is a perfect 3.0—the maximum possible score with the scoring options that were selected—because it corresponds to the alignment of `train02` onto itself.

3. Place the reference ligand in the workspace by clicking the appropriate In box in the Alignments table.

   `train02` is overlaid onto the hypothesis. The features of the hypothesis are perfectly positioned on the matching sites of this ligand, because the hypothesis comes from the reference ligand, so its features must coincide with those of the reference ligand.

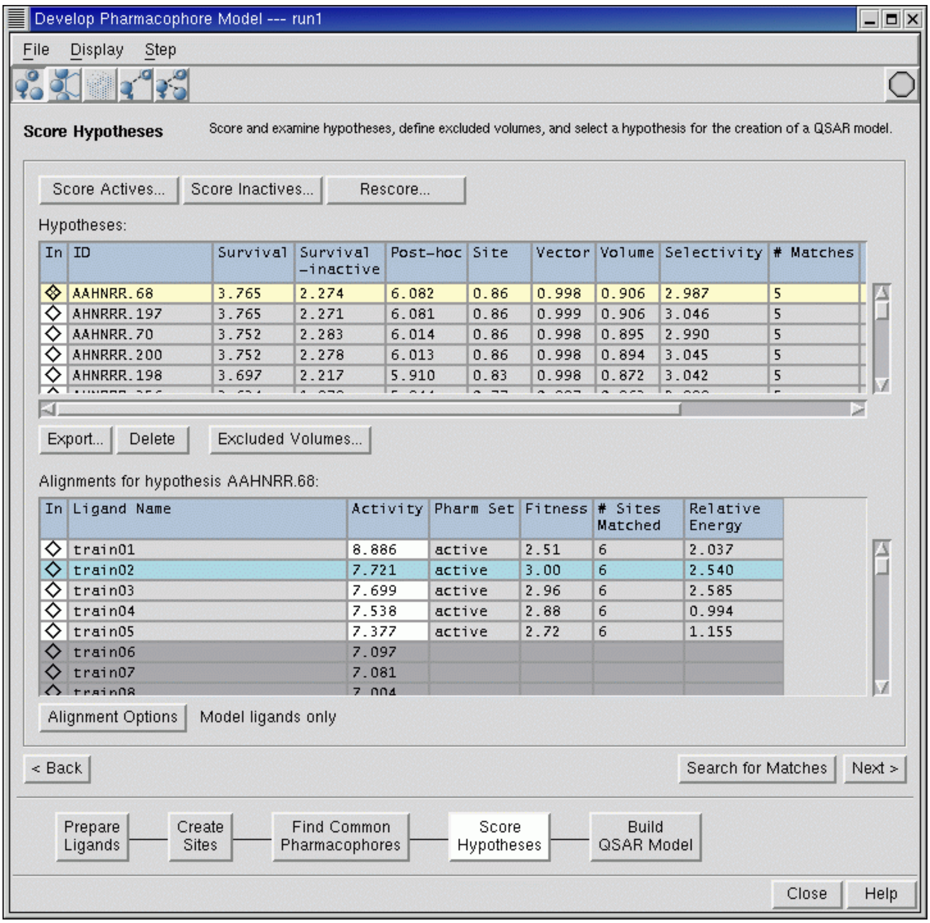


Figure 3.12.** **The** Score Hypotheses **step after scoring.

4. Include `train03` in the Workspace by control-clicking its In box.

   `train03` is overlaid onto the hypothesis by aligning its 6-point pharmacophore with the six features in the hypothesis. Observe that the `train03` and `train02` ligands are only slightly out of alignment with each other.

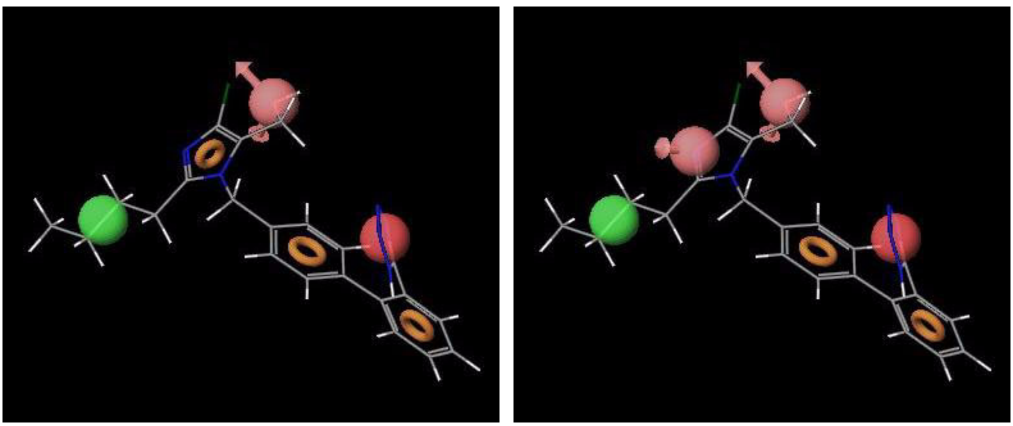
5. Include the remaining actives in the Workspace.

   Most ligands superimpose on each other with only minimal offset. This should come as no surprise because most of the Fitness scores are close to 3.0. In fact, ligands `train02`, `train03`, and `train04` only differ by the substitution of Cl or I for H. The exception is `train01`, whose fitness score is about 2.5. If you place this ligand in the Workspace with the reference ligand, you can see the deviation in the alignment.

6. Examine the second highest scoring hypothesis by selecting the second row of the Hypotheses table.

   The Alignments table is updated to reflect the best alignments for `AHNRRR.197`. Once again, `train02` is the reference ligand. The side-by-side view of `train02` aligned to each of the two highest scoring hypotheses is shown in Figure 3.13.

   It should be evident that the same reference ligand conformation was selected in both cases, and that the two hypotheses differ only in whether the imidazole acts as an aromatic feature or as an acceptor. In situations like this, it is difficult to assess which hypothesis is most reasonable, so one often pursues both for further investigation. For this tutorial, we shall examine both `AHNRRR.197` and `AAHNRR.68`.



`AHNRRR.197` `AAHNRR.68`

Figure 3.13. Hypotheses `AHNRRR.197` ***(L) and*** `AAHNRR.68` ***(R) with reference ligand*** `train02`***.***

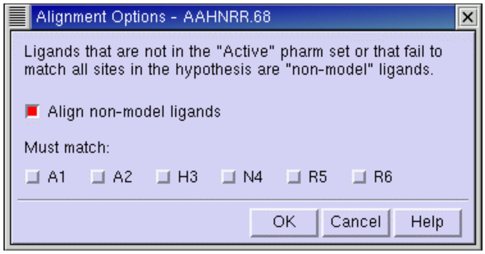


Figure 3.14. The Alignment Options ***dialog box.***

In the next step you will build a 3D QSAR model, which requires alignments for ligands outside the active set. The 49 training and test set ligands cannot all be expected to match all six sites in the hypothesis. Some of the weak binders will be missing one or more features contained in the hypothesis. To deal with this possibility, Phase uses *partial matching* to obtain alignments for these ligands. If at least three sites in the hypothesis can be matched, an unambiguous alignment is obtained. For each ligand outside the active set, then, Phase searches for matches involving the largest possible number of sites, and identifies the match that yields the highest fitness score. Alignments can give information about which features are important and which are not, especially for actives that are not in the training set.

7. Select `AAHNRR.68` in the hypothesis table.
8. Click Alignment Options below the Alignments table.

   The Alignment Options dialog box opens. In addition to selecting the option to align the "non-model" ligands (those that were not in the pharm set or did not match all sites in the hypothesis), you can require certain features to match when the alignment is performed. If the features do not match for a particular ligand, the ligand is not aligned. For this tutorial we will not require any matches.

9. Select Align non-model ligands, and click OK.

   After a short delay, the Alignments table contains an entry for all ligands. It is not always possible to obtain an alignment for every ligand, but it has happened in this case. Scrolling down the table, you will find matches involving 4, 5, and 6 sites and a wide range of fitness values.

Now you will examine the alignments for several ligands: `train12`, which matches all 6 points but has a poor fitness score, `train25`, which matches 4 points and is inactive, and `test02`, which only matches 4 point but is highly active.

10. Place `train12` in the Workspace.

The features in `train12` do not align well with the corresponding features of the hypothesis. This does not necessarily mean that `train12` could never achieve better alignment, but rather that the conformations we are working with contain no such alignment. However, an examination of the structure shows that a better alignment is unlikely, and there is some internal strain in achieving the current alignment: the energy of this conformer relative to the lowest is more than 4 kcal/mol. When building a QSAR model, it is a good idea to take into account the quality of the superimpositions, especially for members of the training set. Cleaner superimpositions usually yield models with greater statistical significance and greater predictive ability.

11. Place `train25` in the Workspace.

    It is immediately obvious that, although the four features that do match align well, there are two features missing. It is reasonable to suppose that at least one of these features is critical to activity.

12. Place `test02` in the Workspace.

    This is a highly active compound in the test set. It also only matches on four features, but there are two features in the molecule that might match if the tolerances were a little looser. This kind of variation implies that there could be some flexibility in the receptor that accommodates some variation in the location of the pharmacophore features.

3.23 Proceeding to Build QSAR Model

Phase QSAR models are based on partial least-squares (PLS) regression, applied to a large set of binary-valued variables that encode whether or not ligand atoms or ligand features occupy various cube-shaped elements of space. Using the hypotheses `AHNRRR.197` and `AAHNRR.68`, you will develop QSAR models to explain the activity data (`pIC50-Exp`), then apply these models to make activity predictions for the test set ligands. Although these hypotheses are the top-scoring hypotheses, there is no necessary connection between the score and the quality of the QSAR model. When you build QSAR models, you should try several hypotheses to ensure that you have a good model. Phase permits you to build models for multiple hypotheses simultaneously.

1. Select `AHNRRR.197` and `AAHNRR.68` in the Hypotheses table.
2. Click either Build QSAR Model or Next.

   The Build QSAR Model step is displayed.

3.24 Assigning Training and Test Set Memberships

By default, all ligands are placed in the training set, so you must separate them into the proper training and test sets. You need only do this for one hypothesis, because the set membership is the same for all hypotheses in this step.

1. Select all the test set ligands (`test01`, `test02`, etc.) in the Alignments table.

   You can do this by clicking `test01` and shift-clicking `test25`. Make sure you do not click in the QSAR Set column.

2. Control-click the QSAR Set column for one of the selected ligands, to change the value from `training` to `test`.

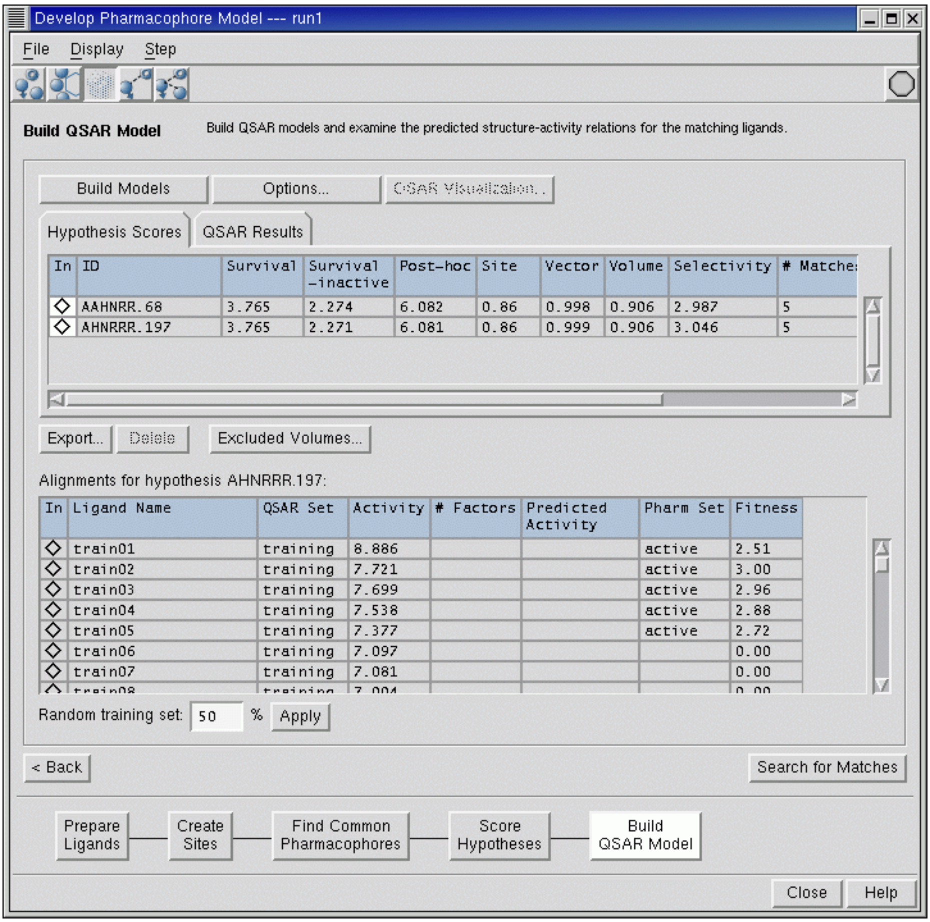


Figure 3.15. The initial view of the Build QSAR Model ***step.***

This results in a training set of 25 ligands and a test set of 24 ligands. If you wanted to eliminate a particular training set (or test set) ligand from consideration, possibly because of a poor superimposition, you could click the corresponding Set cell until its contents were blank.

Note that the hypothesis is used only to obtain ligand alignments: it does not contribute in any way to the QSAR model itself. But the hypothesis has an *association* with the model, because it defines how ligands should be pre-aligned before applying the model.

3.25 Setting QSAR Model Options

Phase can generate QSAR models that are atom-based or pharmacophore-based. The independent variables in the QSAR model are derived from a regular grid of cubic volume elements that span the space occupied by the training set ligands. Each ligand is represented by a set of bit values (0 or 1) that indicate which volume elements are occupied by either a van der Waals model of the atoms in that ligand or by pharmacophore features of that ligand. In the atom-based model, to distinguish different atom types that occupy the same regions of space, a given cube in the grid may be allocated as many as six bits, accounting for six different classes of atoms. Likewise, the pharmacophore features in the hypotheses are represented by bits for each feature type.

In this exercise, you will be developing an atom-based model, and will set parameters to control the sizes of the cubes and the maximum number of PLS factors to include in the model.

1. In the Develop Pharmacophore Model panel, click Options.

   The Build QSAR Model options dialog box is displayed.

2. Set Grid spacing to 1.0.
3. Set Maximum #PLS Factors to 3.
4. Under Model type, ensure that Atom-based is selected.

   Note that for the pharmacophore-based models, you can adjust the feature radii.

5. Click OK.

   The cubes that define the independent variables will be 1 Å on each side, and atom-based linear regression models will be built containing one, two and three PLS factors.

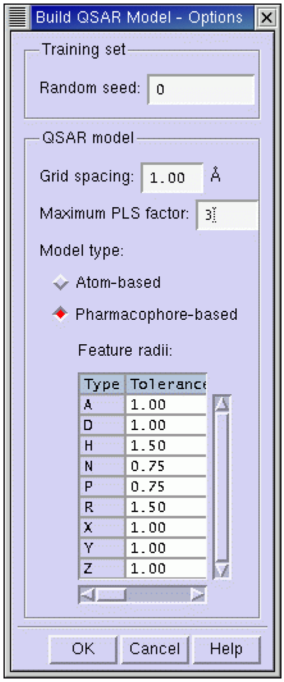


Figure 3.16. The Build QSAR Model - Options ***dialog box.***

3.26 Building the QSAR Models

One-factor, two-factor and three-factor PLS regression models are created using the training set ligands and then applied to the test set ligands.

1. Click Build Models.

   The Start dialog box opens.

2. Select a host, and click Start.

   The octagon at the top right turns green and spins. When it stops, the QSAR results table is filled with data for three PLS regression models. Each regression is itself a QSAR model.

   For a regression with m PLS factors, built using n training set ligands, the statistical parameters shown in the tables are:

SD	The standard deviation of regression. This is the RMS error in the fitted activity values, distributed over *n–m–*1 degrees of freedom.
R-squared	The coefficient of determination. A value of 0.80, for example, means that the model accounts for 80% of the variance in the observed activity data. R-squared is always between 0 and 1.
F	The ratio of the model variance to the observed activity variance. The model variance is distributed over m degrees of freedom and the activity variance is distributed over *n–m–*1 degrees of freedom.
P	The significance level of F when treated as a ratio of Chi-squared distributions. A P value of 0.05 means F is significant at the 95% level.
RMSE	The RMS error in the test set predictions.
Q-squared	Directly analogous to R-squared, but based on the test set predictions. Note that Q-squared can take on negative values if the variance in the errors is larger than the variance in the observed activity values.
Pearson-R	Pearson R value for the correlation between the predicted and observed activity for the test set.

It should be apparent that the R-squared value will always increase as the number of PLS factors increases, but the same is not necessarily true of Q-squared.

3. In the QSAR Results table, click the In column for `AAHNRR.68`.

   The hypothesis is displayed in the Workspace, and the Alignments table is populated with data for this hypothesis, including activity predictions for each ligand using one, two and three PLS factors.

   Scrolling down to the test set, we see that some predictions are good and some are bad. As might be expected, the prediction for `test02`, the active ligand that matched only four sites in the hypothesis, is significantly lower than experiment.

   Of particular interest is `test17`, whose experimental activity value may be in question since its structure was identical to that of `test24` (see Section 3.5 on page 31). The 2-factor prediction for this ligand is 4.37, while the experimental activities originally reported for test17 and test24 were 5.770 and 4.553, respectively. The first experimental value, 5.770, was retained by Phase, so the error in the prediction is quite large. To see how much this observation is skewing the statistics for the test set, you can remove it and rebuild the model.

4. In the Alignments table, click the QSAR Set cell for `test17` to change its contents from `test` to an empty cell.

   A dialog box appears warning you that you will invalidate data in forward steps.

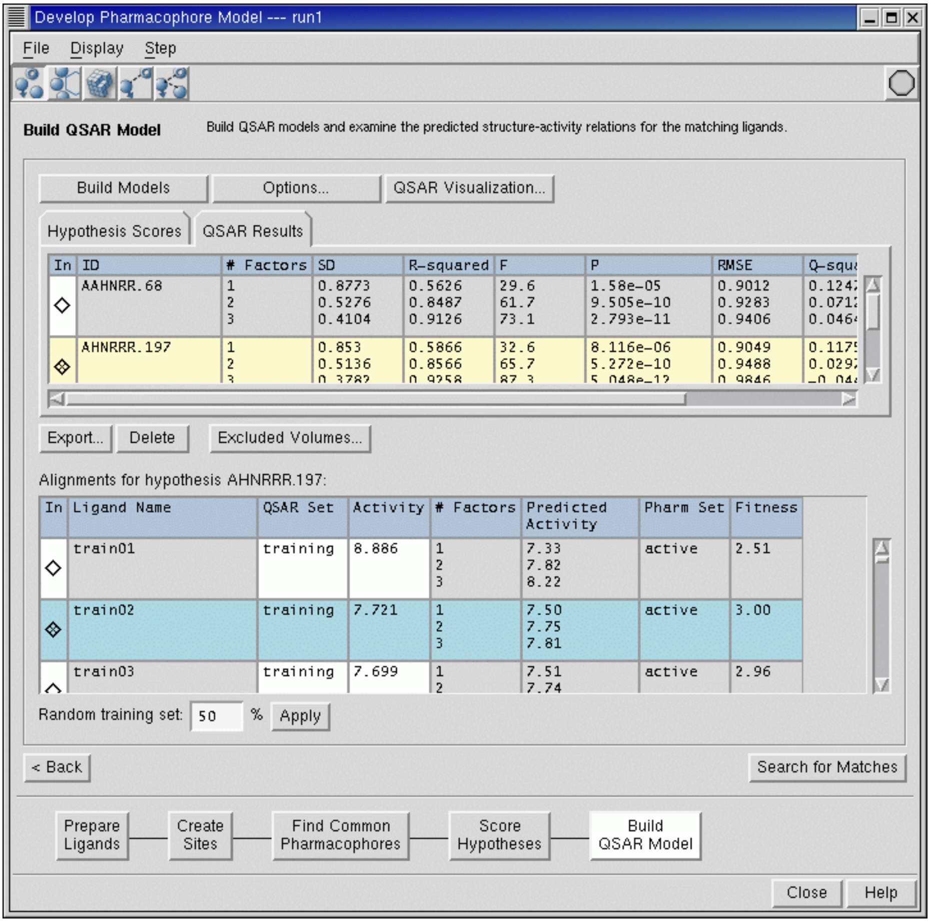


Figure 3.17. The Build QSAR Model ***step after building models.***

5. Choose the Proceed option and click OK.

   The ligand `test17` is no longer a member of the test set.

6. Click Build Models.

   The new models are identical to the old models because the training set is the same. However, the Q-squared value for the test set has increased because the poor prediction for `test17` is not being considered.

3.27 Visualizing the QSAR Model

Three-dimensional aspects of the QSAR model are examined to help gain an understanding of how the structures of the ligands contribute to the computed activity.

1. Ensure that the View QSAR model toolbar button is selected.

2. Place `train01`, the most active ligand, in the Workspace.
3. Click QSAR Visualization.

   The QSAR Visualization Settings panel is displayed. This panel has various options for displaying characteristics of the QSAR model.

   You can view the cubic volume elements occupied by the ligands in the Workspace, or view all the cubes in the QSAR model (i.e., the union of cubes occupied by the 25 training set ligands). You can also view the cubes associated with all atom classes or a specific atom class. The number of PLS factors determines which of the three regression models you are viewing, and the positive and negative coefficient thresholds allow you to see only the cubes whose PLS regression coefficients exceed a particular tolerance.

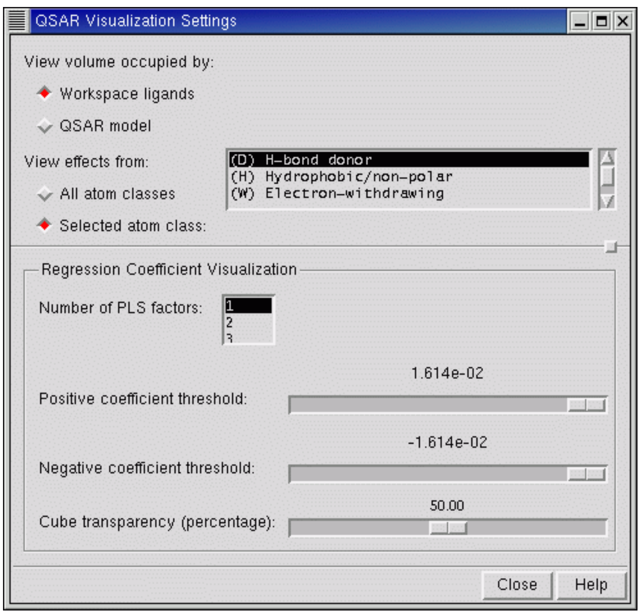


Figure 3.18. The QSAR Visualization Settings ***panel.***

4. Ensure that Workspace ligand and All classes are selected.
5. Select 2 PLS factors.
6. Move the positive and negative coefficient threshold sliders to an intermediate value, such as 0.020 and –0.020.

   In the Workspace, you will see many blue cubes and a small number of red cubes. The blue cubes indicate regions that are favorable for activity and the red cubes indicate regions that are unfavorable. Note that you are viewing only those cubes whose regression coefficients exceed the intermediate thresholds we set.
7. Change the positive and negative coefficient thresholds to more extreme values, such as 0.035 and –0.035.

   The number of cubes in the Workspace drops significantly because now you are viewing only very significant terms in the model.
8. Change the thresholds back to 0.020 and –0.020.
9. Remove `train01` from the Workspace and replaced it with `train25`, the least active ligand.

   The Workspace now contains many more red cubes than blue cubes, indicating a preponderance of unfavorable interactions.
10. Add `train01` back into the Workspace, while keeping `train25` there as well (control-click the In column for `train01`).

    You are now viewing the union of the cubes occupied by the two ligands. It should be evident that `train25` fails to occupy much of the favorable blue volume of `train01`. It should also be evident that this volume is associated primarily with the aromatic and negative features that are not matched by `train25`. So while the hypothesis does not contribute in an explicit sense to the QSAR model, it is reflected implicitly in the regression coefficients and volume occupation patterns. Note that a Phase QSAR model may involve regions of space that extend beyond the physical bounds of the hypothesis because the QSAR considers the volume occupied by all atoms in the ligands.
11. Remove `train25` from the Workspace (click `train01`).

The current view of the model illustrates effects from all atom classes simultaneously, but it is possible to separate out the contributions of individual atom classes.

12. Under View effects from, select Selected atom class.

    By default you will be viewing (D) H-bond donor, which includes polar hydrogens bonded to nitrogen, oxygen and sulfur.

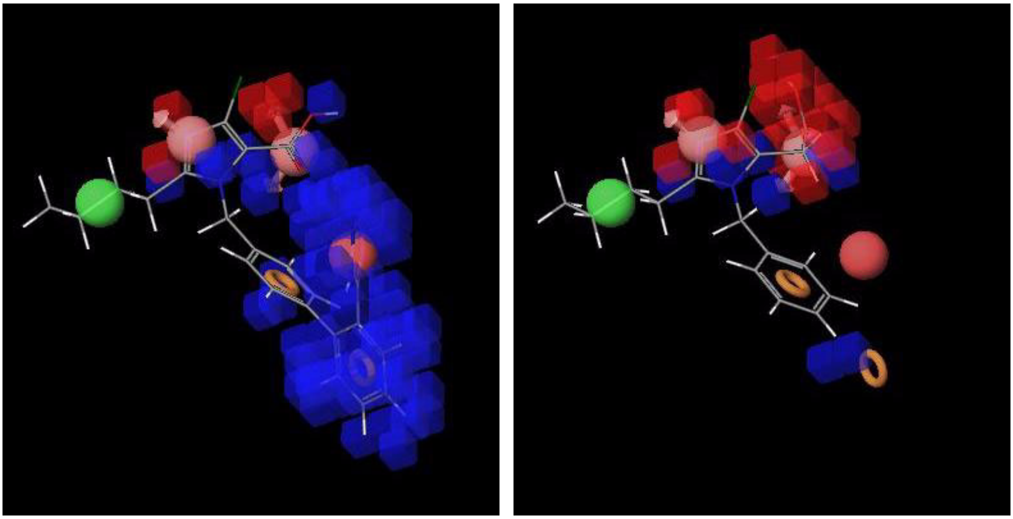


Figure 3.19. QSAR model for ligands* `train01` *and* `train25`*.

13. Change the positive and negative coefficient thresholds to lower values, such as 0.004 and –0.004.

    The view in the workspace indicates that the tetrazole and carboxylic acid groups of `train01` occupy D-type volume, and that this occupation contributes favorably to activity. While we know that both of these groups are likely to be ionized, the QSAR model takes each ligand structure *as is* and assigns each atom to a particular class. So these groups would give rise to negative ionic atoms (type “N”) only if the ligand structures were explicitly ionized.

14. Change the Selected atom class to (H) Hydrophobic/non-polar.

15. Change the positive and negative coefficient thresholds to 0.015 and –0.015.

    Blue cubes are now visible throughout much of the structure of `train01`. These are favorable regions occupied by carbons, halogens, and nonpolar hydrogens. Note that the QSAR model does not distinguish between aromatic and nonaromatic carbons.

16. Add `train25` back into the Workspace.

    Many red cubes appear, indicating that `train25` occupies a fair amount of unfavorable volume of type H.

17. Change the Selected atom class to `(W) Electron-withdrawing`.

    The Workspace indicates favorable and unfavorable regions occupied by electron-withdrawing nitrogen and oxygen atoms. W-type volume includes hydrogen bond acceptor atoms, but it does not distinguish cases where a lone pair may not be available due to conjugation, e.g., amide nitrogens.

Experiment with different atom classes and different thresholds to view the various ways in which the QSAR model distinguishes ligands with high and low activities.

Creating a 3D Database

Once a pharmacophore model has been developed, it may be used to search a database, with the goal of identifying additional active molecules. This chapter describes the process of creating Phase 3D databases. You can search a set of 3D structures in a structure file, but if you plan to search the same database with the same feature set multiple times, it is quicker to create a Phase database that includes conformer sets and sites.

If you plan to search the database on a host that is different from the one you will use in this chapter, you must create the database on a file system that is accessible to the other host.

4.1 Creating the Database Directory

The Phase database is a set of files that resides in a directory. We refer to this directory as the *database* directory. You can create the database directory yourself, or Maestro can create it for you. In this exercise, you will create the database directory yourself.

1. In a terminal window, navigate to your `phase_tutorial/databases` directory.

   This directory must be part of a file system with at least 100 MB of available disk space.

2. Create a new directory named `testDB` by entering the following command:

   ```
   mkdir testDB
   ```

4.2 Creating a New 3D Database

In this exercise you will create a small database from a Maestro file that contains single-conformer models of 100 druglike molecules. The conformers and sites will be stored in the database.

1. Start Maestro in the database directory, `phase_tutorial/databases/testDB`.

   If Maestro is already running, change to the database directory by entering the following command in the Commands text box in the main window:

   `cd` *database-directory*

   This command operates the same as the UNIX command, and changes the directory relative to the Maestro current working directory.

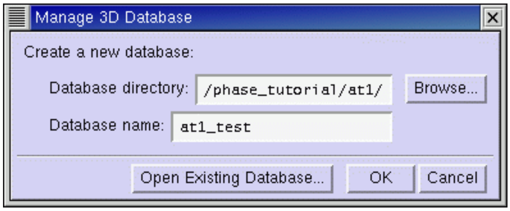


Figure 4.1. The New 3D Database ***dialog box.***

2. Choose Manage 3D Database from the Phase submenu of the Applications menu in the main window.

   If this is the first time that you have created a Phase database, the New 3D Database dialog box is displayed. If you have previously created a database, the Manage 3D Database panel is displayed, with the most recent database you used already loaded. In the latter case, choose New from the Database menu to display the New 3D Database dialog box.

3. Click the Browse button to the right of the Database directory text box.

   A directory selector opens. If you started Maestro in the database directory, this directory should be displayed in the Selection text box. If it is not, navigate to the directory and select it.

4. Click OK.

5. Enter `testDB` in the Database name text box.

6. Click OK.

   The Manage 3D Database panel opens, or if it is already open, the Structures table is cleared.

4.3 Adding Structures to the Database

You can add structures to a database from multiple sources. These sources can contain single molecules, or conformer sets, and both can be stored in the database. When you add structures from a file, you can choose to clean up the structures if they are 2D or are not well optimized.

1. In the Manage 3D Database panel, click Add.

   The Add Structures dialog box opens.

2. For Add from, select File.

3. Click Browse.

4. Navigate to the directory `phase_tutorial/databases/userFiles`

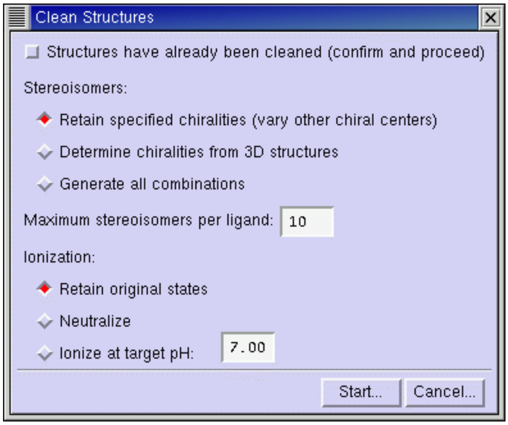


Figure 4.2. The Clean Structures ***dialog box.***

5. Select `dbMolecules.mae` and click OK.
6. Under File contains, ensure that Individual structures is selected.

   This file contains single molecules, not conformer sets.
7. Click OK.

   The Clean Structures dialog box opens.

Databases are often created from chemical files that contain fairly crude structures, so it is important to clean the structures unless you are certain that you have good 3D models with the proper numbers of hydrogens attached. If you are certain that you have good structures, you can select Structures have already been cleaned, and click Start. The job that is run simply performs some checks on the structures. In this exercise, you will clean the structures.

8. Ensure that Structures have already been cleaned is not selected.
9. Under Stereoisomers, select Retain specified chiralities.

   This option creates stereoisomers only for structures that contain chiral centers of unspecified chirality.
10. Under Ionization, select Retain original states.

    In practice, you may want to consider the Neutralize or the Ionize option. These options are useful if you want to create databases with consistent ionization states, and guarantee that pharmacophore features are mapped consistently from one database to the next.
11. Click Start in the Clean Structures dialog box, and again in the green Start dialog box.

This job takes 5–10 minutes on a 2 GHz Pentium 4 processor. Upon completion, the structures are added to the Structures table. The database contains 180 molecules. The number of molecules has increased from 100 to 180 due to the creation of stereoisomers. You are now ready to generate conformations. Before you do this, however, you will import feature definitions for site creation, which is done automatically after conformer generation. In the next exercise, you will export these cleaned structures to the Project Table.

4.4 Exporting Structures to the Project Table and a File

In this exercise, you will export the cleaned structures from the database to the Project Table, and then to a file, so they can be used externally. You will re-import these structures later on, so that the database contains both entries with conformers and entries without conformers.

1. Ensure that all the structures in the table are selected.

   To do this, right-click in the table and choose Select All from the shortcut menu.

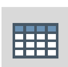
2. Click Export.

   The Export Structures dialog box opens.

3. Click OK.

   It does not matter which option is selected in the dialog box, since there is only one conformer per entry at this stage. The structures are added to the Project Table.

4. Click the Open/Close Project Table button on the main toolbar.

   The Project Table panel opens.

5. Select the structures that were exported.

   If you started this chapter with a new Maestro session, this should be all structures in the Project Table.

6. From the Export submenu of the Table menu, choose Structures.

   The Export dialog box opens.

7. Ensure that Maestro format is selected and the source of structures is the selected entries.

8. Navigate to the directory directory `phase_tutorial/databases/userFiles`.

9. Add `dbClean.mae` to the end of the path in the File text box, and click OK.

   The file `dbClean.mae` is created.

4.5 Importing Feature Definitions

The feature definitions determine how pharmacophore elements are mapped to each conformer, and therefore the types and locations of pharmacophore sites in the database. It is important to remember that each hypothesis used to search a database also has an associated set of feature definitions, and that a search cannot be performed unless the two sets of definitions are identical. So if you create a database using a particular set of feature definitions, and you wish to search it with a hypothesis created using a different set of definitions, then you will either have to create new sites and store them in the database, or generate sites while the database is searched. Either option is considerably more expensive than performing a search in which the definitions are identical at the outset.

It is therefore advisable to adopt a single set of feature definitions that will be used for database creation and pharmacophore hypothesis development. This is not always possible, of course, because you may need to modify the definitions for a particular pharmacophore model project, or, as time passes, you may identify new definitions that you want to incorporate into all of your future projects and databases. But if you work toward a stable set of feature definitions, you will spend less time searching and updating 3D databases.

By default, the database uses a set of default or built-in feature definitions. If, as discussed above, you have a particular set of 'universal' definitions that you have adopted, then you would import them here. In this exercise you will import a set of feature definitions.

1. Choose Edit Features from the Database menu.

   The Edit Features dialog box opens.

2. Click Import From File.

   A file selector labeled Import Feature Definitions opens.

3. Navigate to the directory `phase_tutorial/databases/userFiles`

4. Select `phase_feature.def` and click Import.

   The file selector closes, and the definitions in the table should be updated to reflect what was imported from the file.

5. Click OK

   The Edit Features dialog box closes, and the Create Sites dialog box is displayed. This happens because the features have been changed, and the database might have sites that were created with the previous feature set. In this case no sites are needed because they will be automatically created when the conformers are generated.

6. Ensure that Only structures that have sites already is selected, and click OK.

   Since the structures you imported don't have sites yet, the site creation job will do nothing. The job is run, to check the database for invalid sites.

You are now ready to generate conformers and create sites for each structure in the database.

4.6 Generating Conformers

A 3D database will typically be searched using hypotheses developed from various sets of ligands that bind to distinct biological targets. Naturally, then, the database must provide reasonable coverage of conformational space to increase the chances of finding active conformations against any given target.

1. Ensure that all the structures in the table are selected.

   To do this, right-click in the table and choose Select All from the shortcut menu.

2. Click Generate Conformers.

   The Generate Conformers dialog box is displayed.

3. Under Generate conformers for, select Selected structures.
4. For Sampling, ensure that Rapid is selected.
5. Ensure that Maximum number of conformations per ligand is 100.

   There is only one search method available from the MacroModel search method option menu, Ligand torsion search. You can use other search methods outside Phase and import the conformer sets.

6. Ensure that Preprocess is not selected.
7. Ensure that Postprocess is not selected.

   The generated conformations will be somewhat crude, but reasonable coverage will still be achieved, and with minimal computational expense. MacroModel postprocessing is only recommended for very small databases, or when the database is being created simply to obtain high quality alignments for a set of ligands that will be used, for example, to build a QSAR model using the Phase command line interface.

8. Click Start in the Generate Conformers dialog box, and again in the Start dialog box.

This job takes 15 minutes on a 2 GHz Pentium 4 processor. Upon completion, you should see the Conformers column in the Structures table filled out with the numbers of conformers generated. For many molecules, the maximum of 100 confomers was reached. After the conformers are generated, a second job is run to create sites for the conformers.

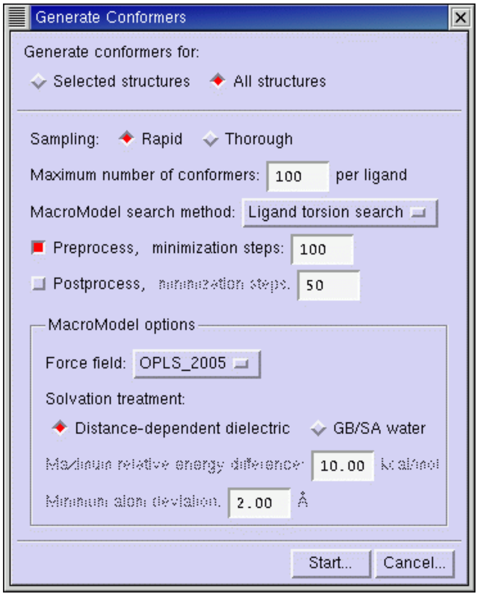


Figure 4.3. The Generate Conformers ***dialog box.***

Pharmacophore sites are now created and the database can be searched. At this point you can proceed to the database searching tutorial in Chapter 5, or continue with the optional exercises.

4.7 Creating Subsets

There may be occasions in which you will not want to search every molecule in a database. For this reason, Phase supports the concept of database subsets, which are essentially lists of database entries. When a search is conducted, it is possible to search against the entire database or against only a subset of entries that you define in the Subsets panel. In this exercise you will create two subsets, one of structures that have conformers and sites, and one of structures that do not have conformers and sites.

1. Ensure that all the structures in the table are selected.

   To do this, right-click in the table and choose Select All from the shortcut menu.

2. From the Database menu, choose Subsets.

   The Subsets panel opens.

3. Click New.

   The New Subset dialog box opens.

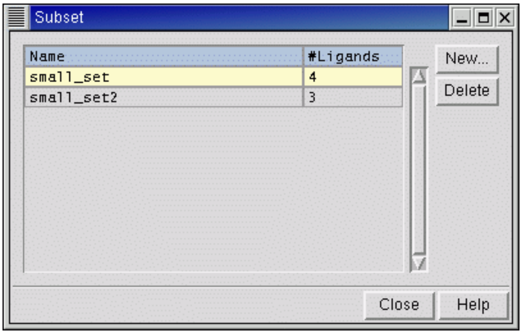


Figure 4.4. The Subset ***panel.***

4. Enter `conformers` in the Subset name text box.
5. Ensure that Currently selected structures is selected, and click OK.

   The subset is listed in the table in the Subsets panel.

For the next subset, you will import the structures that you exported in Section 4.4, and use these to define a subset of structures that does not have conformers.

6. In the Manage 3D Database panel, click Add.

   The Add Structures dialog box opens.
7. For Add from, select File, then click Browse.
8. Navigate to the directory `phase_tutorial/databases/userFiles`
9. Select `dbClean.mae` and click OK.
10. Under File contains, ensure that Individual structures is selected, and click OK.

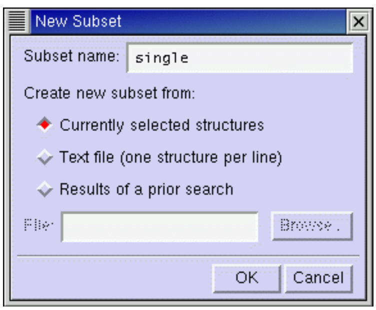


Figure 4.5. The New Subset ***dialog box.***

11. In the Clean Structures dialog box, select Structures have alredy been cleaned.

    These are the structures that you cleaned before, so there is no need to clean them again. However, a job is still run to check for invalid structures.

12. Click Start in the Clean Structures dialog box, and again in the Start dialog box.

    When the job finishes, the structures are listed in the Structures table of the Manage 3D Database panel.

13. Select the newly added structures.

    These structures should have a 1 in the Conformers column. Click the first, shift-click the last to select them.

14. In the Subset panel, click New.

    The New Subset dialog box opens.

15. Enter `single` in the Subset name text box.

16. Ensure that Currently selected structures is selected, and click OK.

    The subset is listed in the table in the Subsets panel.

You have now created two subsets, one that contains structures with conformers and sites, and one that contains single structures without sites. These two sets will be used in the database search in the next chapter. You may now close the Subsets panel and the Manage 3D Database panel, and quit Maestro or proceed to the database search exercises in Chapter 5.

Finding Matches to a Hypothesis

One of the primary reasons for developing a pharmacophore model is to accelerate the identification of new active compounds. This is most commonly done by searching a 3D database for matches to a pharmacophore hypothesis. In this chapter, we demonstrate how an existing pharmacophore hypothesis is used to search the 3D databases created in Chapter 4. If you have not already completed Chapter 4, you must do so before you can begin this portion of the tutorial.

5.1 Preparing for the Exercises

1. Start Maestro in the `phase_tutorial` directory (see Section 4.1).

   If Maestro is already running, change to this directory by entering the following command in the Commands text box in the main window:

   `cd` *tutorial-directory*

   Use the full path for *tutorial-directory*.

2. Choose Find Matches to Hypothesis from the Phase submenu of the Applications menu in the main window.

   The Find Matches to Hypothesis panel opens, with an empty Available hypotheses table and various options for matching and hit treatment.

Phase database searching is normally broken into two steps: finding and fetching. In the find step, which is the most expensive part of the search, the database is scanned for geometric arrangements of pharmacophore sites that match the hypothesis within a tolerance applied to the intersite distances. A brief summary of each set of matching points is written to a match file, which is subsequently used in the relatively inexpensive fetch step as a lookup table to retrieve hits from the database.

A hit is nothing more than the matching conformation after it has been aligned to the hypothesis using a standard least-squares technique applied to the site points. Hits are returned in order of decreasing fitness, which is a user-adjustable measure of the quality of the alignment of the hit with respect to the reference ligand conformation from which the hypothesis was derived:

$$S = W_{\text{site}}(1 - S_{\text{align}}/C_{\text{align}}) + W_{\text{vec}}\,S_{\text{vec}} + W_{\text{vol}}\,S_{\text{vol}}\,.$$

Table 5.1. Description of parameters in the fitness scoring function.

Parameter	Description
S_{align}	Alignment score: RMS deviation between the site point positions in the matching conformation and the site point positions in the hypothesis.
C_{align}	Alignment cutoff. User-adjustable parameter; default is 1.2.
W_{site}	Weight of site score. User-adjustable parameter; default is 1.0.
S_{vec}	Vector score: average cosine between vector features in the matching conformation and the vector features in the reference conformation.
W_{vec}	Weight of vector score. User-adjustable parameter; default is 1.0
S_{vol}	Volume score: Ratio of the common volume occupied by the matching conformer and the reference conformer, to the total volume (the volume occupied by both). Volumes are computed using van der Waals models of all non-hydrogen atoms.
W_{vol}	Weight of volume score. User-adjustable parameter; default is 1.0

The terms in the score are described in Table 5.1.

As hits are fetched from the database, they can be filtered using excluded volumes, and their activities can be predicted using a QSAR model. You can also control the maximum total number of hits and the maximum number of hits per molecule.

The rationale for separating the steps of finding and fetching is that you might want to explore different settings associated with fitness, excluded volumes, and so on, without having to repeat the more expensive find step. So there are Matching options, which are associated with the find step, and Hit treatment options, which are applied only when fetching.

5.2 Choosing the Database and Hypothesis

Before performing a search, you must select a database and a hypothesis. If you have not used another Phase database since you generated the `testDB` database in the last chapter, this database is already selected.

If the testDB database is not selected:

1. Click Browse.

   A file selector opens, with a filter of `*_phase.inp`.

2. Navigate to the `phase_tutorial/databases/testDB` directory.

3. Select the file `testDB_phasedb` and click OK.

   The path to the database is now listed in the File name text box.

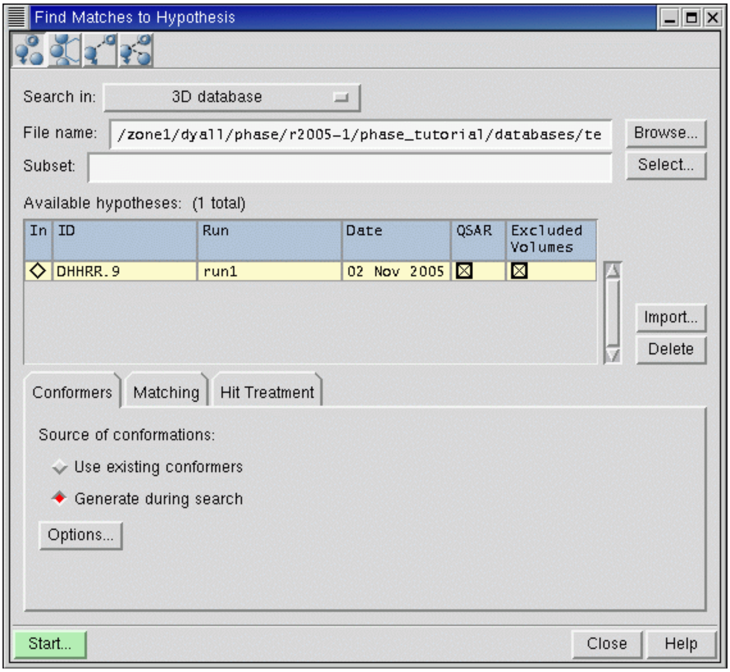


Figure 5.1. The Find Matches to Hypotheses ***panel.***

For the first database search, you will use the subset that contains conformers and sites.

1. Click Select.
2. In the Select Subset dialog box, select `conformers` and click OK.

   The subset name is listed in the Subset text box.

Next, you will import the hypothesis. If you click Search for Matches in the Develop Pharmacophore Model panel, the selected hypotheses are carried forward into the Available hypotheses table. Likewise, if you edit hypotheses in the Edit Hypotheses panel, these are also available for database searching.

3. Click Import.

   A file selector labeled Import Hypothesis opens.

4. Navigate to the directory `phase_tutorial/databases/userFiles`.
5. Select `DHHRR_9.tab` and click OK.

The hypothesis `DHHRR_9` is listed in the Available hypotheses table. The QSAR and Excluded Volumes boxes are selected, indicating that the hypothesis has both QSAR models and excluded volumes associated with it.

5.3 Performing a Standard Search

In this exercise, you will perform an ordinary search involving both finding and fetching, using default options.

1. Ensure that Use existing conformers is selected in the Conformers folder.

   The options in the other folders can be left at the defaults.

2. Click Start.

   The Start dialog box opens. In this dialog box you can enter a job name, select the host to run the job and the number of processors to use. For the exercises in this chapter, you will run the job on the local host.

3. Enter `conformers` in the Name text box.

4. Click Start in the Start dialog box.

   The Monitor panel opens, allowing you to view the progress of the search. You will see output from the find step followed by output from the fetch step. The entire search should take less than a minute, after which a total of 48 hits are placed in the Project Table.

5. Click the Open/Close project table button on the main toolbar.

   The Project Table panel is displayed. The hits are in an entry group that is named after the job, i.e. `conformers`. You may view the hits in the Workspace and examine the variety of properties that are written to the Project Table, including the matching sites indices, the fitness and its various components, and the predicted activities from the QSAR model.

Roughly a quarter of the database molecules produced hits. This is quite a high fraction, and indicates that this particular hypothesis is not very selective. Generally speaking, hypotheses with greater numbers of sites, a variety of features, and widely spaced site points are most selective because they are less likely to be matched within a general population of molecules.

5.4 Searching Against Existing Matches

In this exercise you will use the matches found in the previous section, but with different hit treatment, i.e., you will rerun only the fetch step.

1. In the Matching folder, select Use existing matches.
2. In the Hit treatment folder, deselect Apply excluded volumes.
3. In the Hit treatment section, reduce the hits total setting from 1000 to 10.
4. Click Start.
5. In the Name text box of the Start dialog box, enter `noexclvol`.
6. Click Start.

   The Monitor panel shows output from only the fetch step. When the job is finished, a total of ten hits appear in the Project Table, in an entry group named `noexclvol`.

5.5 Searching with Site and Conformer Creation

In this exercise you will search the database subset for which conformers and sites were not generated. The conformers and sites will be generated during the search.

1. In the Find Matches to Hypothesis panel, click Select.
2. Select the subset `single` and click OK.
3. In the Conformers folder, select Generate during search.
4. In the Matching folder, select Find new matches.
5. In the Hit treatment folder, select Apply excluded volumes and return the hits total setting to 1000.
6. Click Start.
7. In the Name text box of the Start dialog box, enter `single`.
8. Click Start.

The output to the Monitor panel now includes information about the conformation generation. This job should take several minutes, because the conformers are being generated during the search. When the job is finished, a total of 48 hits appear in the Project Table, in an entry group named `single`. The results should be identical to those obtained in the first search.

Getting Help

Schrödinger software is distributed with documentation in PDF format. If the documentation is not installed in `$SCHRODINGER/docs` on a computer that you have access to, you should install it or ask your system administrator to install it.

For help installing and setting up licenses for Schrödinger software and installing documentation, see the *Installation Guide*. For information on running jobs, see the *Job Control Guide*.

Maestro has automatic, context-sensitive help (Auto-Help and Balloon Help, or tooltips), and an online help system. To get help, follow the steps below.

- Check the Auto-Help text box, which is located at the foot of the main window. If help is available for the task you are performing, it is automatically displayed there. Auto-Help contains a single line of information. For more detailed information, use the online help.
- If you want information about a GUI element, such as a button or option, there may be Balloon Help for the item. Pause the cursor over the element. If the Balloon Help does not appear, check that Show Balloon Help is selected in the Help menu of the main window. If there is Balloon Help for the element, it appears within a few seconds.
- For information about a panel or the folder that is displayed in a panel, click the Help button in the panel. The Help panel is opened and a relevant help topic is displayed.
- For other information in the online help, open the Help panel and locate the topic by searching or by category. You can open the Help panel by choosing Help from the Help menu on the main menu bar or by pressing CTRL+H.

  To view a list of all available Phase–related help topics, choose Phase from the Categories menu in the Categories tab. Double-click a topic title to view the topic.

If you do not find the information you need in the Maestro help system, check the following sources:

- *Maestro User Manual*, for detailed information on using Maestro
- *Maestro Command Reference Manual* for information on Maestro commands
- *Phase User Manual*, for detailed information on using Phase
- Frequently Asked Questions pages, at https://www.schrodinger.com/Phase_FAQ.html

The manuals are also available in PDF format from the Schrödinger Support Center. Information on additions and corrections to the manuals is available from this web page.

If you have questions that are not answered from any of the above sources, contact Schrödinger using the information below.

E-mail:	help@schrodinger.com
USPS:	Schrödinger, 101 SW Main Street, Suite 1300, Portland, OR 97204
Phone:	(503) 299-1150
Fax:	(503) 299-4532
WWW:	http://www.schrodinger.com
FTP:	ftp://ftp.schrodinger.com

Generally, e-mail correspondence is best because you can send machine output, if necessary. When sending e-mail messages, please include the following information, most of which can be obtained by entering `$SCHRODINGER/machid` at a command prompt:

- All relevant user input and machine output
- Phase purchaser (company, research institution, or individual)
- Primary Phase user
- Computer platform type
- Operating system with version number
- Phase version number
- Maestro version number
- mmshare version number

Glossary

Active compound—A compound that shows high affinity for the biological target. Synonymous with the term *ligand*.

Active set—The set of active compounds that is used to develop a pharmacophore model. This set does not necessarily include all active compounds.

Excluded volume—A region of space in a pharmacophore hypothesis that should not be occupied by any atom of an active compound.

Feature—see **Pharmacophore feature**

Hit—A structure in a 3D database that is found to contain an arrangement of site points that can be mapped to the pharmacophore hypothesis. A hit is not necessarily active, but it is presumed to have a greater than average probability of being active if it was retrieved using a valid hypothesis.

Hypothesis—see ***n*-Point pharmacophore hypothesis**

Inactive compound—A compound that shows little or no affinity for the biological target.

Intersite distance—The distance between any two site points in a pharmacophore.

Ligand—see **Active compound**

Negative compound—A compound that is inactive, yet highly similar in structure to one or more known actives. Some compounds are negative because they lack certain key pharmacophore features found in true actives. Other negatives may actually satisfy exactly the same pharmacophore hypotheses as the actives, but possess extraneous structural characteristics that prevent binding.

Pharmacophore feature—A characteristic of chemical structure that may facilitate a noncovalent interaction between a ligand and a biological target. Examples are hydrogen-bond acceptor ("A"), hydrogen-bond donor ("D"), hydrophobe ("H"), positive ionic center ("P"), negative ionic center ("N").

Pharmacophore site—The labeling and location of a particular pharmacophore feature within a molecule. For example, a hydrogen bond acceptor site could simply be a nitrogen atom that carries an available lone pair. A hydrophobic site might be a methyl carbon or the centroid of a phenyl ring. The term *site point* is often used interchangeably with pharmacophore site.

Pharm set—The set of all compounds, active and inactive, used to develop a pharmacophore model.

***n*-Point pharmacophore**—Any 3D arrangement of *n* pharmacophore features.

***n*-Point pharmacophore hypothesis**—A specific 3D arrangement of *n* pharmacophore features, with associated uncertainties in the feature positions. High affinity ligands in their active conformations are expected to contain pharmacophore sites that can be mapped (within the limits of uncertainty) to any valid hypothesis. A given hypothesis may contain features that are associated with a single mode of binding, or it may contain features that are common to two or more modes of binding.

Reference ligand—The ligand that provides the pharmacophore that defines a hypothesis. In pharmacophore model development, this pharmacophore yields the highest multi-ligand alignment score for the active-set ligands. The reference ligand matches the hypothesis exactly, and has a perfect fitness score.

Site point—see **Pharmacophore site**

3D Database—A set of molecules, each of which is represented by one or more 3D conformational models, augmented with a pharmacophore-based representation of the molecules. A 3D database includes feature types and site point coordinates for each conformation.

Variant—The set of feature types in a pharmacophore. For example, the variant AHH indicates a 3-point pharmacophore containing one hydrogen bond acceptor and two hydrophobic sites.

Vector feature—A pharmacophore feature that contains directionality, such as a hydrogen bond acceptor, hydrogen bond donor, or aromatic ring. A vector feature does not necessarily have vector geometry.

Vector geometry—the geometric characteristics of hydrogen bond acceptors and donors. Refers to the direction of lone pairs in a hydrogen-bond acceptor or the direction of the heavy-atom–hydrogen-atom bond in hydrogen-bond donors. Features with vector geometry must be vector features.

120 West 45th Street
32nd Floor
New York, NY 10036

101 SW Main Street
Suite 1300
Portland, OR 97204

3655 Nobel Drive
Suite 430
San Diego, CA 92122

Dynamostraβe 13
68165 Mannheim
Germany

QuatroHouse, Frimley Road
Camberley GU16 7ER
United Kingdom

SCHRÖDINGER®